THE IMPACT OF TYPE 2 DIABETES-RELATED COMPLICATIONS ON UTILITY AND HEALTHCARE COSTS, AND SELF-REPORTED HEALTH RELATED QUALITY OF LIFE AS A PREDICTOR OF MORTALITY IN DIABETES

D.PHIL. THESIS

OXFORD, TRINITY TERM 2013

Maria L. Alva

Division of Public Health

Wolfson College

University of Oxford

ACKNOWLEDGEMENTS

My thanks go to my supervisor, Professor Alastair Gray, whose encouragement, patience and guidance from the initial to the final phase enabled me to complete this thesis and divulgate its results.

I wish to offer thanks to my examiners, Dr. Ramon Luengo-Fernandez and Professor Paul Fenn, for their comments.

I owe thanks to the Diabetes Trial Unit, headed by Rury Holman, who allowed me to use the UK Prospective Diabetes Study (UKPDS) data, product of 30 laborious years of research. In particular I would like to thank the data administrators, Ian Kennedy and Ruth Coleman, who patiently answered my many questions.

My heartfelt gratitude also goes to Dr. Borislava (Boby) Mihaylova for her advice, acumen and invaluable frankness; Dr. Jose Leal for acting as a mediator in the task of collecting Hospital Episode Statistics (HES); Leicester Gill and Mathew Davidson from the Unit of Health-Care Epidemiology at Oxford University for facilitating the linkage between English HES and UKPDS.

I would also like to thank the staff at HERC for their camaraderie. My office mates for brightening many lunch-times. The Wolfson rowing team for keeping me fit and grounded for 2 years.

Last but not least, I wish to thank my family: my parents, Armando Alva and Germana Chiola, who supported me continuously through my many endeavours; Robert and Judy Zeitlin, the best in-laws a person could have; my husband, Andrew Zeitlin, who encouraged me lovingly in the many ups and downs that accompany the life of a researcher; my daughter, Noemi, who was born between chapters 3 and 4 and refuelled my desire to succeed. To her I dedicate this thesis.

PUBLICATIONS AND PRESENTATIONS

PUBLICATIONS

Alva M., Gray A., Mihaylova B., Clarke P. (2013), "The Effect of Diabetes Complications on Health-related Quality of life: the importance of longitudinal data to address patient heterogeneity" *Health Economics.*

CONFERENCE PRESENTATIONS

1. 24th June 2010: **Health Economist's Study Group** (HESG), Cork, Ireland. "Estimating the quality of life impact of diabetes related complications: new results from the UKPDS".
2. 18th September 2010: **The Mount Hood Five Challenge**, University of Lund, Malmo, Sweden. "Estimating the quality of life impact of diabetes related complications: new results from the UKPDS".
3. 23rd September 2010: **European Association for the Study of Diabetes**, 46th Annual Meeting, Stockholm. "Estimating the quality of life impact of diabetes related complications: new results from the UKPDS".
4. 26th June 2011: **American Diabetes Association**, 71st Annual Meeting, San Diego, USA. "Outpatient resource utilization of type-2 diabetes patients: New evidence from the UKPDS"
5. 11th July 2011: **iHEA** 8th World Congress, Toronto, Canada, "The Effect of Diabetes Related Complications on the Marginal Utility: estimating the bias due to unobservable heterogeneity using the UKPDS".
6. 4th October 2011: **Clinical Trials Methodology Conference**, Bristol, UK, "The Effect of Diabetes Related Complications on the Marginal Utility: estimating the bias due to unobservable heterogeneity using the UKPDS".
7. 7th June 2012: **Mt Hood Challenge**, John Hopkins University, Baltimore, USA, "Inpatient and Outpatient Resource use in Patients with type-2 Diabetes: New Evidence from the UKPDS"
8. 27th June 2012: **Health Economist's Study Group** (HESG), Oxford UK, "Inpatient and non-inpatient resource utilization of type-2 diabetes patients"

SEMINAR PRESENTATIONS

1. 11th of February 2010: **Public Health Work in Progress Seminar**, Oxford "The Impact of Diabetes-Related Complications on Hospital-in-patient Costs".
2. 2nd June 2010: **HERC Research Discussion meetings**, Oxford. "Estimating the quality of life impact of diabetes related complications over time".
3. 3rd of March 2011: **Public Health Work in Progress Seminar**, Oxford "Outpatient Resource use in Patients with type-2 Diabetes".
4. 12th December 2011: **HERC Research Discussion meetings**, "Is Self-reported quality of life an independent predictor of Survival?" Oxford.

ABSTRACT

THE IMPACT OF TYPE 2 DIABETES-RELATED COMPLICATIONS ON UTILITY AND HEALTHCARE COSTS, AND SELF-REPORTED HEALTH RELATED QUALITY OF LIFE AS A PREDICTOR OF MORTALITY IN DIABETES

Maria L. Alva
Wolfson College, University of Oxford
D.Phil. Thesis
Trinity Term 2013

Background: This thesis focuses on the economic analyses of type-2 diabetes complications defined as macro-vascular (myocardial infarction, stroke, ischemic heart disease, heart failure) and micro-vascular (amputation and eye-related complications leading to blindness in one eye). Diabetes-related complications are a substantial component of the overall economic, physical and psychological burden of the disease. As the efforts in treating diabetes are geared towards reducing the likelihood of complications, understanding the welfare benefits and future savings from reducing diabetes complications is paramount in determining the cost-effectiveness of competing diabetes therapies.
Aims: The thesis is divided into three essays aiming to (1) characterize changes in the health-related quality of life of diabetes patients over time and assess the contributions of diabetes complications to these changes; (2) study the drivers of healthcare expenditure for people with diabetes in terms of both inpatient care and non-inpatient resource utilization, and estimate the impacts of diabetes-related complications on health care costs; (3) understand the role played by self-reported quality of life in predicting mortality after controlling for clinical risk factors.
Methods: This thesis uses longitudinal data to answer the questions of interest. A unifying theme across the thesis is the challenge of estimating causal parameters in a context in which there may be substantial observed and unobserved patient heterogeneity.
Findings: Failing to account for patient heterogeneity, and in particular un-measurable variation in patients' outcomes, is likely to bias the impact of complications on quality of life and on non-inpatient costs, as well as to confound predicted time to death. In the case of QoL, ignoring heterogeneity is likely to overestimate the impact of complications on self-reported utility because the patients who will eventually experience diabetes-related complications are already on a lower utility path compared to those who do not. In the case of both inpatient and non-inpatient costs, patients who go on to develop complications have higher cost both pre and post complications. In the case of inpatient costs there is no evidence that unobserved patient heterogeneity matters, while in the case of non-inpatient utilization the hypothesis of a common baseline level of utilization is rejected in the subset of patients that contribute to the FE identification. This subset however is systematically different from the sample as a whole, being predominately more likely to have complications and other causes of hospitalization. Moreover, a trade-off occurs when we are interested in predictions; models that exploit within-patient variation have wider confidence intervals and have thus less precision than population average models. The final substantive chapter finds that HRQoL is significantly associated with survival at the population level and that when patient specific unobserved heterogeneity is taken into account, the power of QoL to predict life expectancy increases. Neglected heterogeneity in frailty causes underestimation of both the extent of positive duration dependence and the impacts of time varying covariates.

TABLE OF CONTENTS

CHAPTER 3:

INPATIENT AND NON-INPATIENT COSTS

CHAPTER 4:

LIFE EXPECTANCY

CHAPTER 5:

CONCLUSION

LIST OF TABLES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

APPENDICES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

LIST OF COMMON ABBREVIATIONS

CI confidence interval

DAG directed acyclic graph

ER emergency room

EQ-5D EuroQol-5 dimensions

FE fixed effects

GEE generalized estimating equations

GLM generalized linear model

GP general practitioner

HR hazard ratio

HRQoL health-related quality of life

IHD ischemic heart disease

IVMR inverse Mill's ratio

MI myocardial Infarction

OLS ordinary least squares

PA population average

PH proportional hazard

PMS post monitoring study

QALY quality-adjusted life year

QoL quality of life (also referred to as HRQoL, health-related QoL)

RCT randomised controlled trial

RE random effects

SD standard deviation

SE standard error

SG standard gamble

T2DM diabetes mellitus type 2

TD therapy decision

TTO time-trade-off

Chapter 1

Chapter 1 provides a conceptual overview of this dissertation and its objectives. Section 1.1 outlines the research questions addressed, and the original contributions of the thesis. Section 1.2 presents an overview of the disease area and in particular what it means to have diabetes in terms of co-morbidities. This is followed by a brief overview in Section 1.3 of the United Kingdom Prospective Diabetes Study (UKPDS) and the UKPDS Post Trial Monitoring (PTM) study.

1.1 Objectives and Structure of the thesis

The objectives of the thesis are threefold:

1. To characterize changes in the health-related quality of life of patients with type-2 diabetes over time and to assess the significant determinants and predictors of these changes. These are the topics of chapter 2.
2. To study the drivers of healthcare expenditure, in particular the impacts of diabetes-related complications, in terms of both inpatient and non-inpatient health care utilization and costs. This is the subject of chapters 3.
3. To understand the role played by self-reported quality of life in predicting mortality after controlling for clinical and demographic risk factors. This is studied in chapter 4.

These objectives are important in the context of establishing the cost effectiveness of diabetes interventions, especially as the range of therapies to prevent Type 2 diabetes-related complications is rapidly expanding. Diabetes-related complications lead to reduced

quality of life, significant cost to the healthcare system, and reduced life expectancy. As welfare benefits and future savings from reducing diabetes complications may have a large effect on the economic evaluation of diabetes therapies, it is important to have accurate estimates of the impacts of diabetes-related complications on patients' quality of life and healthcare costs, as well as of the impact of quality of life on survival.

The contributions of the thesis include the exploration of appropriate panel data techniques, the importance of functional form issues, the implications of self-selection in questionnaire response, and clustering and persistence in questionnaire response when attempting to answer the questions of interest. The unifying theme across chapters 2, 3, and 4 is the challenge of estimating causal parameters in a context in which there may be substantial unobserved patient heterogeneity.

Some of the research questions at which this dissertation is aimed, particularly those regarding quality of life (QoL) and resource utilization in the trial population, have been addressed by past UKPDS studies (Clarke, Gray et al. 2002; Clarke, Gray et al. 2004). Those studies consisted of cross-sectional analyses. My current work benefits from having access to a substantial amount of additional data, including several rounds of post-trial questionnaire responses; this is rare, especially in the literature that analyses the impact of complications on wellbeing. Multiple rounds of questionnaires on healthcare utilization and QoL were collected for the purpose of disentangling the effects of acute events (short-term costs accruing in a relatively short time-frame) from *state* costs (long term costs). This set up allows for longitudinal analyses which present three major advantages over conventional cross-section analyses:

1 A larger number of data points increases the degrees of freedom of the estimations, reducing the collinearity among binary explanatory variables (i.e., whether or not a patient has a complication, when the complications are rare), which improves the efficiency of the estimates.

2 The possibility to investigate potential biases arising from survey non-response over time. Respondents may have intrinsically different characteristics than non-respondents, so that the answers of the former may systematically differ from the potential answers of non-respondents.

3 The ability to control for the effects of missing or unobservable time-invariant confounding factors that may otherwise bias results. Purging the correlation of these confounding factors with the covariates of interest is necessary, albeit not sufficient, for the estimation of causal parameters.

1.2 Type 2 diabetes

Diabetes mellitus[1] type 2 (referred to more succinctly as type 2 diabetes or T2DM) is a disorder characterized by chronic hyperglycaemia resulting from defects in insulin secretion, insulin action, or both (Smushkin and Vella 2010). Although competing theories exist, from weight loss management to surgical intervention, the condition is generally agreed to be non-reversible and therefore the treatment and management strategy aims to enable patients to lead lives free of complications.

The World Health Organization predicts that by 2030 diabetes will affect more than 366 million individuals worldwide (Wild, Roglic et al. 2004). According to *Diabetes UK*, the

[1] Mellitus means "sweet as honey" *(lat.).* The term refers to the excess sugar found in the urine of diabetic patients.

known diagnosed population in the United Kingdom in 2010 is 2.8 million. The prevalence rates reported in Table 1.1 refer to those currently diagnosed with either type 1 or type 2 diabetes. By 2025 it is estimated that over four million people will have diabetes in the UK. Most of these cases will have Type 2 diabetes, owing to the UK's ageing population and rapidly rising numbers of overweight and obese people.

Table 1.1: Prevalence of Diabetes in the UK (2010)

Country	Prevalence	Number of people
England	5.40%	2,338,813
Northern Ireland	3.70%	68,980
Scotland	4.10%	223,943
Wales	4.90%	153,175

UK average = 4.26 %. Source: Diabets UK.

According to the YHPHO (Yorkshire and Humber Public Health observatory), around 92% of people with diabetes in England have the type-2 variety (Table 1.2).

Table 1.2: Distribution of diabetic types 1 and 2 in the population

	Type 1	Type 2
Male	10%	90%
Female	5%	95%
All	8%	92%

Source: (2005 Diabetes Population Prevalence model)

Despite the fact that type-2 diabetes is more common than type-1, there are large ethnic and geographic variations in the prevalence of the two forms. For example, In Scandinavia, where type-1 diabetes is more common, type-2 diabetes is thought to account for about 85% of all cases. In Asia, only a few percent of the population have type-1 diabetes while the majority have the type-2 form (Groop 2005).

Type-2 diabetes appears to be more common in people who are overweight and of African, Afro-Caribbean and Asian ethnicity. The risk of developing diabetes is around 20

times greater[2] in the very obese (those with a body mass index of over 35) than in people with a body mass index of between 18 and 25. At the population level, people from minority ethnic communities are reported to have up to six times higher than average risk of developing diabetes (PBS, Diabetes Prevalence Model (2005)). Ethnicity, however, is often a catch-all variable. The interesting question to ask would be, what omitted behavioural factors contribute to the strong association between race and outcomes? This is a question that the dataset used is not geared to answer but should be of paramount concern in terms of healthcare planning and community education. In the absence of any genetic factors that may create race based predispositions to diabetes, other variables need to be studied.

Compared with Type-1, Type-2 diabetes usually develops in older populations (>40 years). This feature may however change. A few decades ago the reported presence of type-2 in children was considered an anomaly; recent trends indicate it is no longer such a rarity with a significant increase in the number of diagnosed cases. Conservative estimates suggest there are around 100 children with T2DM in the UK. Some researchers (Lobstein and Leach 2004) state that the figure could be as high as 1,400 based on the surveyed number of overweight and obese school children in the UK.

Diabetes is one of the most demanding chronic conditions in terms of the variety of potential co-morbidities that are directly associated with it. Compared with the non-diabetic population, people with diabetes have a higher prevalence of micro and macro vascular complications[3] (Morgan, McEwan et al. 2006). This leads to an increase in hospitalizations and outpatient costs and to a reduced health utility for the individual.

[2] http://www.dh.gov.uk/

[3] Diabetes complications are generally separated into macrovascular complications (coronary artery disease, peripheral arterial disease, and stroke) and microvascular complications (diabetic nephropathy, neuropathy, and retinopathy).

1.3 UKPDS

My thesis employs data from the UK Prospective Diabetes Study (UKPDS). The UKPDS was a landmark randomized multicentre trial of glycaemic therapies in 5,102 patients with newly diagnosed type-2 diabetes. These subjects had no existing serious complication at the time of diagnosis.

Trial Structure and Description

The UKPDS trial ran for twenty years (1977 to 1997) in 23 UK clinical sites. Between 1977 and 1991, 5,102 subjects were recruited (Table 1.3). The flow of patients in the study during the initial randomization is shown in Appendix A1.1.1. The patients had a 3-month dietary run-in during which they attended monthly a UKPDS clinic where they were seen by a physician and dietician. After the run-in, their mean Fasting Plasma Glucose (FPG) level was calculated from an average of three measurements over two weeks. Patients with FPG levels above and below a pre-established range (6-15mmol) were excluded from the trial. Non-overweight patients were randomly assigned to either intensive treatment (2,279 patients) or conventional treatment with diet (1,138 patients). If marked hyperglycaemia or other symptoms were found, patients in the conventional treatment arm were secondarily randomised to treatment with sulphonylurea or insulin therapy, with the additional option in a separate stratified randomisation of metformin administered to overweight patients. In subsequent analyses, the initial 342 patients allocated to metformin were compared with 411 similarly overweight patients in the conventional arm. In 1987, 1,148 hypertensive patients (mean age 56, mean blood pressure at entry 160/94 mm Hg) were allocated to either tight control of blood pressure (758 patients) or to less tight control (390 patients) for

a median follow up of 8.4 years (for a schematic representation of this second trial please refer to Figure A 1.2 in the Appendix).

Table 1.3: The 23 UKPDS participating centres

Cities	Number of Participants	Year	
		Recruitment Start	End
Aberdeen	225	1982	1989
Belfast City	167	1980	1989
Belfast Victoria	294	1983	1989
Birmingham	349	1979	1989
Carshalton	241	1982	1989
Derby	130	1988	1991
Dundee	213	1982	1989
Hammersmith	207	1979	1988
Ipswich	249	1982	1989
Leicester	82	1988	1991
London	364	1978	1989
Manchester	102	1988	1991
Northampton	178	1987	1991
Norwich	280	1982	1989
Oxford	443	1978	1989
Peterborough	98	1987	1990
Salford	170	1988	1991
Scarborough	130	1987	1991
St. George's	441	1979	1989
Stevenage	173	1983	1988
Stoke on Trent	205	1983	1989
Torbay	137	1987	1991
Whittington	224	1982	1989
Total	**5102**		

The final results, published in 1998, showed for the first time that the complications of T2DM could be reduced by improving blood glucose and blood pressure control (UKPDS(33) 1998, UKPDS(38)1999).

From September 1997 the intervention phase of the study closed and surviving patients were treated to best practice by their GPs. At that point all patients entered a post-trial monitoring study, which continued to collect information on clinical events. The

administration of resource use and quality of life questionnaires initiated at the end of 1996 continued until September 2007.

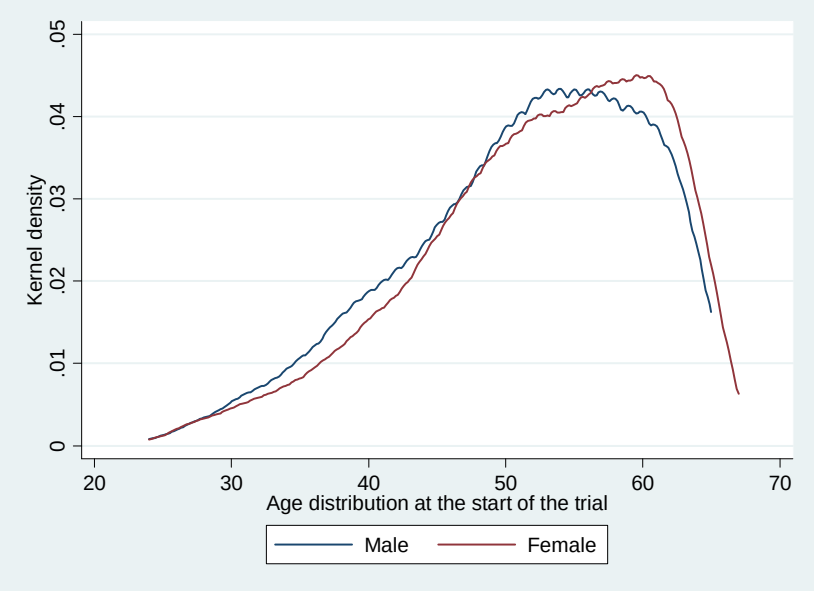
In my study, information on randomisation is not exploited because the main focus is the post-trial monitoring (PMT) period during which all patients were offered best standard of care. The scope of my study is the analysis of Quality of Life (QoL) and resource utilization post-trial, when patients were monitored by their general practitioners according to best practice. It is important however to acknowledge that if treatments were correlated with complications but treatments also affected QoL directly, through side-effects for example, then there is a potential for omitted variable bias and hence the need in an ideal world to control for treatment effects. The value of a randomized controlled trial (RCT) is in explaining the impact of a treatment (TX) on an outcome, which could be complications, QoL impacts, costs, etc. However an RCT is like a black box in the sense that it may be difficult to explore intermediary effects, especially if these are in the causal path of interest. For instance, if we are interested in the impact of a TX on QoL, we cannot control for complications, as the impact of treatments on QoL will be biased towards zero, assuming that the TX decreases complications, which in turn increases QoL. In other words, complications enter the causal path of the effect of TX on QoL. Thus we do not learn anything from the impact of TX on QoL by adding complications or any covariates that affect the treatment path. Only if we assume that QoL changes are obtained purely by decreasing complications and if there are no side effects or other mechanisms that may influence QoL, can we use RCT results to say something about the value of additional/alternative therapies. If we want to talk about the impact of complications on QoL it is better to study these associations directly. Since one cannot randomize complications, panel methods like the ones discussed in this thesis give credibility to this

kind of analysis. The results that lie ahead could be used as input into a structural approach to modelling the economic benefits of alternative treatments.

Characteristics of Subjects in the UKPDS

The average age at diagnosis of participants in the UKPDS trial was 52.4. The youngest and the oldest participants at diagnosis were 24 and 67 years old, respectively. Women at the start of the trial were on average a year older than men (Figure 1.1). 58.96% of the UKPDS sample is composed of men. Thus gender in the trial is not representative of the actual distribution of diabetes across genders as the prevalence of T2DM in the population tends to be higher for women than for men (Wild S, Roglic G et al. 2004). Self-selection into trials is often unavoidable and the unintended cause of this is that in general women and ethnic minorities tend to be underrepresented despite the effort to achieve a balanced demographic representation.

Figure 1.1: Age Distribution at recruitment in the UKPDS trial



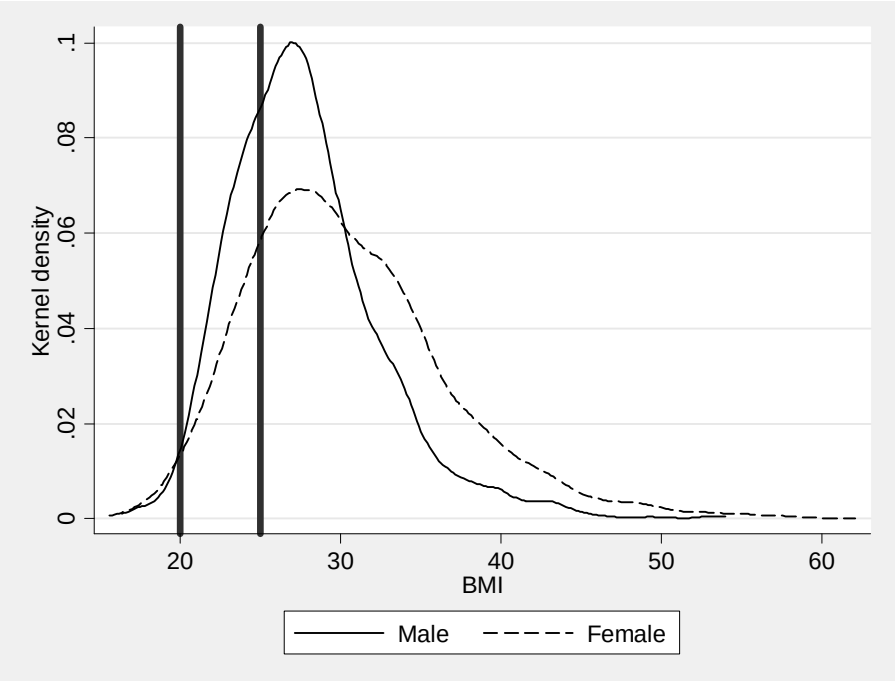
The ethnic composition of the UKPDS patients (Table 1.4) however was broadly in line with the 1999 Health Survey for England (Gray, Clarke et al. 2002).

Table 1.4: Ethnic distribution in the UKPDS trial

Ethnicity	Freq.	Percent
White Caucasian	4,180	81.9%
Asian Indian	387	7.6%
Afro-Caribbean	496	9.7%
Asian Chinese	12	0.2%
Other	27	0.5%
Total	5,102	100

Figure 1.2 shows the distribution of body mass index (BMI) for males and females in the UKPDS. BMI was calculated by dividing weight in kilograms by height in metres squared. The vertical lines show the thresholds (20-25) within which healthy BMIs should lie. Women entering the trial have on average a 9.4% higher BMI than their male counterparts. Both sexes however are on average overweight (BMI is on average 27.8 and 30.4 for males and females respectively). The average person in the UK has a BMI of 25.4(Wang, McPherson et al. 2011).

Figure 1.2: BMI distribution for males and females at the start of the trial



A number of biometric statistics and behavioural variables such as BMI, smoking status, exercise and alcohol consumption were either collected in the trial only at baseline, or at subsequent rounds preceding the PTM period and thus the administration of questionnaires.

Data on levels of activity, alcohol consumption and smoking status were collected at triennial intervals up to 2002. The number of observations available however decreases sharply with the passage of time. Information on activity levels is available for 5,075 patients. A person is defined as sedentary if he or she not only rarely exercises but also has a sedentary job. A moderately active person is somebody who reportedly gardens or plays golf at weekends but otherwise does very little of any other form of exercise. An active person is defined as someone who stays on their feet much of the day, or walks or cycles to work more than one mile, or gardens or plays golf 3 times a week or more[4]. Finally a person is defined as fit if their occupation is manual and/or they engage in vigorous exercise three or more times a week. We have 23,005 person-years of data, less than 4.5% of which involve vigorous exercise. The *Between* column in Table 1.5 reports the breakdown of activity levels in terms of patients rather than person-years. It indicates that during the period of the study 2,572 patients remained sedentary, 3,784 stayed active and so on, for a grand total of 10,555 patients falling into any of the four categories. However, we only have information on 5,075 patients, which is explained by the fact that patients switch across categories over time. The *Within* percent measures the overall stability for any given level of activity as it represents the fraction of time a person has a specific value. The total *Within* of 48.08% is the normalized value between weighted average of the *Within* percents. The transition matrix explains and expands on this concept.

[4] The wording on activity types and what they entail was provided by the UKPDS

In Table 1.6 the bold diagonal row reflects the initial values, and the columns reflect the final values. Each year, 55.95% percent of patients remained sedentary in the next year; the remaining 30.96% became moderately active, 12.08% became active and approximately 1% went on to be categorized as fit.

Table 1.5: Tabulation of Activity levels across all patients

N=5075	**Overall**		**Between**		**Within**
Activity level	Freq.	Percent	Freq.	Percent	Percent
Sedentary	5,065	22.02	2,572	50.68	45.48
Moderately active	8,174	35.53	3,784	74.56	47.63
Active	8,736	37.97	3,556	70.07	53.07
Fit	1,030	4.48	643	12.67	33.58
Total	23,005	100	10,555	207.98	48.08

Table 1.6: Transition matrix of exercise patterns across all patients

Activity Level	sedentary	moderately active	active	fit
Sedentary %	**55.95**	30.96	12.08	1
Moderately active %	21.44	**49.03**	27.88	1.66
Active %	9.24	28.57	**56.99**	5.2
Fit %	4.18	16.73	45.76	**33.33**

For alcohol consumption we have at least one measurement for each of the 5,102 participants, although not all of these are measured at the same time. There were 120 participants who at recruitment had become abstinent or occasional drinkers but who used to be regular or dependent. The majority of participants (76.42%) reported themselves to be occasional drinkers at least once over the maximum of 6 measurements available. Even if in person-year terms only 1.02% of the sample appears to be alcohol dependent, this corresponds to almost 3.3% percent of participants in the sample.

Table 1.7 reports similar information on alcohol consumption.

Table 1.7: Transition matrix of alcohol consumption (%)

Alcohol consumption (N=5102)	none	occasional	regular	dependent
None %	**79.07**	19.94	0.96	0.03
Occasional %	18.08	**75.06**	6.72	0.15
Regular %	4.23	37.5	**54.05**	4.23
Dependent %	5.52	19.02	43.56	**31.9**

It is interesting to note that occasional drinkers have the highest average self reported quality of life. Even regular and dependent drinkers report on average higher levels of QoL than those who abstain. These differences, however, are not statistically significant and may be correlated with a host of other factors such as socio-economic status and ethnicity. 60% of Asian participants, for instance, report never consuming alcohol, compared to less than 20% of Caucasians.

Each year, 99% of those who are non-smokers remain so the following year. The remaining 1% become smokers. Although an individual has a 1% probability of becoming a smoker each year, someone who is currently a smoker has a 6.65% likelihood of becoming (or returning to be) a non-smoker.

Table 1.8: Transition matrix of smoking status (%)

Smoking	no	yes
no	99.08	0.92
yes	6.65	93.35

Columns represent t and rows represent t+1

Biochemical measurements were taken at entry into the trial and thereafter annually up to 2002 for a maximum of 24 years. These are averaged in Table 1.9.

Table 1.9: Biomarkers

Measurements	**Mean**	**SD**	**Min**	**Max**
HbA1c (%)	7.8	1.2	4.7	13.2
SBP (mmHg)	137.2	14.8	87.6	191.1
LDL (mmol/L)	3.4	0.8	0.9	6.7
HDL (mmol/L)	1.1	0.2	0.5	2.6
Albuminuria (milligrams)	69.7	185.2	1.8	2591.0

All the reported averages fall outside the ranges defined as normal in the medical literature. The reference range for HbA1c in a healthy person is about 4%–5.9%; the average in the sample is approximately double that amount. Higher amounts of HbA1c indicate poorer control of blood glucose levels. The average systolic blood pressure (SBP) in the sample is 137.2 mmHg. In adults, the ideal systolic pressure should be less than 120 mmHg. The average low density lipoprotein (LDL) is 3.4 mmol/L. According to the UK Cholesterol Education Program, in the absence of any other risk factors, the optimal LDL level should be less than 2.59 mmol/L. According to the American Heart association guidelines, the optimal levels of fasting High-Density Lipoprotein (HDL) is greater than 1.55 mmol/L while in the sample the average is 1.1. Men tend to have noticeably lower HDL levels than women. Normally, a person secretes less than 25 milligrams of albumin in their urine each day while the average albuminuria secretion in the sample is almost 70 milligrams.

Although not available for all participants, information about the social class of the UKPDS participants was occupation-based (Table 1.10), derived following the HMSO Classification of Occupations (HMSO 1980).

Table 1.10: Socio-economic status of UKPDS participants

Social class	Freq.	Percent
Professional	148	4.51
Managerial	736	22.43
Skilled worker	1,416	43.16
Semi -skilled	772	23.53
Unskilled	209	6.37
Total	3,281	100

Women were coded by their husband's social class, reflecting the social structure of the 1970s and 80s, unless they were widowed, divorced or unmarried in which case their own

occupation was used. Unemployed or retired participants were coded according to their last known occupation. In this classification, skilled workers comprised both manual and non-manual workers. In a published UKPDS paper (Thomason, Biddulph et al. 2005) studying the underlying causes of death in subgroups of diabetic patients, classes were amalgamated into three groups: (1) professional and managerial, (2) skilled, (3) semi-skilled and manual. In terms of wellbeing, patients in the first two socio-economic categories (professional and managerial) have significantly higher QoL than any of the other three groupings despite having the same likelihood of diabetes related complications.

Diabetes Complications

Details of each diabetes-related event were recorded for all patients, and the UKPDS centres concerned were asked to provide full information on these; events were then independently classified into predefined categories based on the 9th revision of the International Classification of Diseases codes (ICD). Clinical results from 10 years of post study monitoring were published in 2008 (Holman, Paul et al. 2008).

The set of micro and macro-vascular complications collected during the trial and in the post trial monitoring phase will be the central independent (explanatory) variables in the subsequent chapters. The underlying conjecture is that diabetes-related complications can be avoided, and in the pursuit of causal parameters to the outcomes of interest (utility, costs and survival) I treat complications as non-random events, the likelihood of which is influenced by risk factors such as blood sugar, blood pressure and lipids.

The endpoints used in the trial have been discussed in various UKPDS studies (Turner RC, Holman RR et al. 1991; Ewart 1999; Adler, Stratton et al. 2000) and were chosen to be major clinical events that affect the life and well-being of patients. The criteria selected by

the UKPDS Data Monitoring and Ethics committee included only those endpoints that had unequivocal external verification, with the exception of four non-fatal diabetes-related endpoints, which were less easy to verify independently and required independent assessors because they involved clinical decisions (i.e. cataract extraction, photocoagulation, vitreous haemorrhage, and heart failure).

Table 1.11: UKPDS Clinical Endpoints

Fatal Endpoints*	Major Non Fatal Endpoints	Minor Non Fatal Endpoints
Vascular Cardiac Death (N=868)	Myocardial infarction (N=633)	Heart Failure (N=284)
Vascular Stroke Death (N=204)	Ischemic heart disease (N=541)	Cataract Extraction (N=182)
Sudden death (N=74)	Stroke (N=389)	Vitreous Haemorrhage (N=284)
Vascular Peripheral Death (N=26)	Blindness in one eye (logMar>1) (N=322)	Retinal Photocoagulation (N=894)
Renal Death (N=57)	Amputation (N=228)	
Hyperglycaemic Death (N=5)	Renal Failure (creatinine>250 μmol/l) (N=76)	
Hypoglycaemic Death (N=6)		

***Other registered causes of death were: accidental death (N=27), cancer related (N=525), other specified cause of death (N=34), and unknown cause of death (N=434).**

The complete list of recorded clinical endpoints together with the number of events registered during the 30 years of the study is illustrated in Table 1.11. *Cardiovascular endpoints* are defined as (i) non-fatal myocardial infarction (International Classification of Diseases, Ninth Revision [ICD-9] code 410); (ii) fatal vascular cardiac event (ICD-9, codes 410-414.9 and 428-428.9); (iii) sudden death (ICD-9, codes 798-798.9); (iv) ischemic heart disease (ICD-9, codes 411- 414.9) and (v) congestive heart failure (ICD-9, codes 428-428.9). These occur two or three times more often in people with Type-2 diabetes than in the

general population and can cause death at an earlier age than is usual (Nichols, Gullion et al. 2004).

Patients with type-2 diabetes also have a higher risk of cerebrovascular disease (i.e. non-fatal (ICD-9, codes 430-434.9, 436) and fatal stroke (ICD-9, codes 430-438.9)). Other complications can also develop, due to closure of small blood vessels. These can cause blindness (herein defined as a visual acuity in one eye with Snellen 6/60 or ETDRS (Early Treatment Diabetic Retinopathy Study) log MAR 1 or worse, persisting for 3 months), kidney failure (defined as creatinine levels of above 250 mmol/l (ICD-9 codes 205.3 and 585-586)) leading to death (ICD-9 codes 580-593.9) and amputations of a digit or limb (ICD-9 codes 5.845-5.848 or 250.6).

3,563 out of 5,102 patients had at least one diabetes related endpoint during the window of observation. 1,538 patients had a major diabetes related complication and 1,535 patients had a minor diabetes related complication (Table 1.11). 642 patients out of the 3,563 (17.84%) had both a major and minor events. The recurrence of complications (Table 1.12) in the sample was low, with the exception of amputations.

Table 1.12: Subjects* experiencing one or more (non-fatal) events

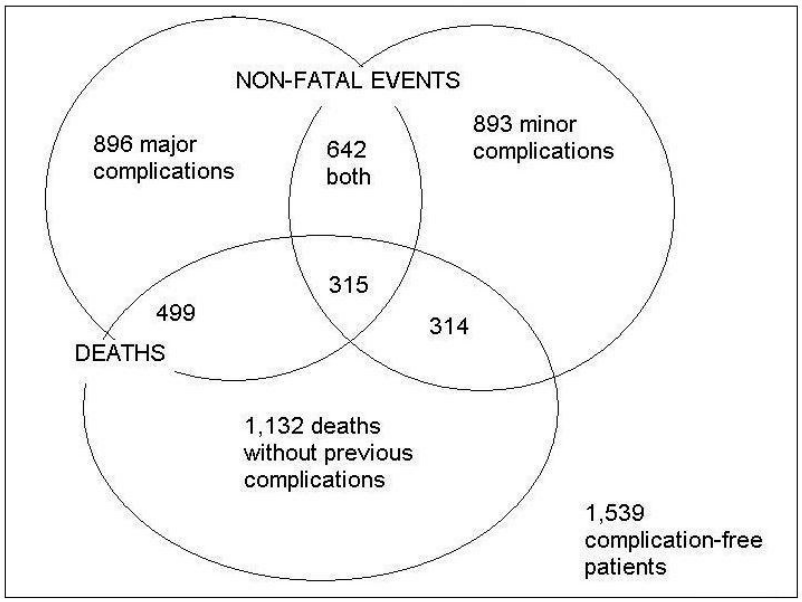
Event	once	twice	thrice	4	5	6	Total subjects	Incidence rate
Myocardial infarction	471	67	8	1			547	10.7%
Ischemic heart disease	521	10					531	10.4%
Stroke	318	31	3				352	6.9%
Blindness in one eye (logMar>1) †	271	24	1				296	5.8%
Amputation	108	33	13	1	1	1	157	3.1%
Renal Failure (creatinine>250 μmol/l)	74	1					75	1.5%
Heart Failure	276	4					280	5.5%
Cataract Extraction	415	392	3				810	15.9%
Vitreous Haemorrhage	109	26	7				142	2.8%
Retinal Photocoagulation	853	20	1				874	17.1%

* The total number of patients enrolled in the trial is 5,102

† defined as persisting for 3 months

2,260 out of 5,102 patients (44.3%) died during the course of the trial. Of these 1,132 patients died without ever experiencing a diabetes related complication. Vascular cardiac deaths (868) alone accounted for more than 38% of all deaths.

Figure 1.3: Set diagram of patient's distribution between fatal and non-fatal events



N=5,102

Questionnaire Administration

A maximum of seven questionnaires per patient were administered with a six-year gap between Questionnaire 1 (administered in 1996/97) and 2 (administered in 2003); subsequent questionnaires were administered annually on the anniversary of the patient's therapy decision (with the exception of the last questionnaire, administered to all patients in October 2007). Patients completed these questionnaires with no help from staff or family during clinic visits. Patients who did not attend clinics during the survey period were sent the questionnaire by mail. A total of 15,434 resource use and quality of life questionnaires were issued to patients alive between 1996/7 and 2007. 12,248 resource use questionnaires were returned (if a patient did not complete one of the resource use fields in a returned

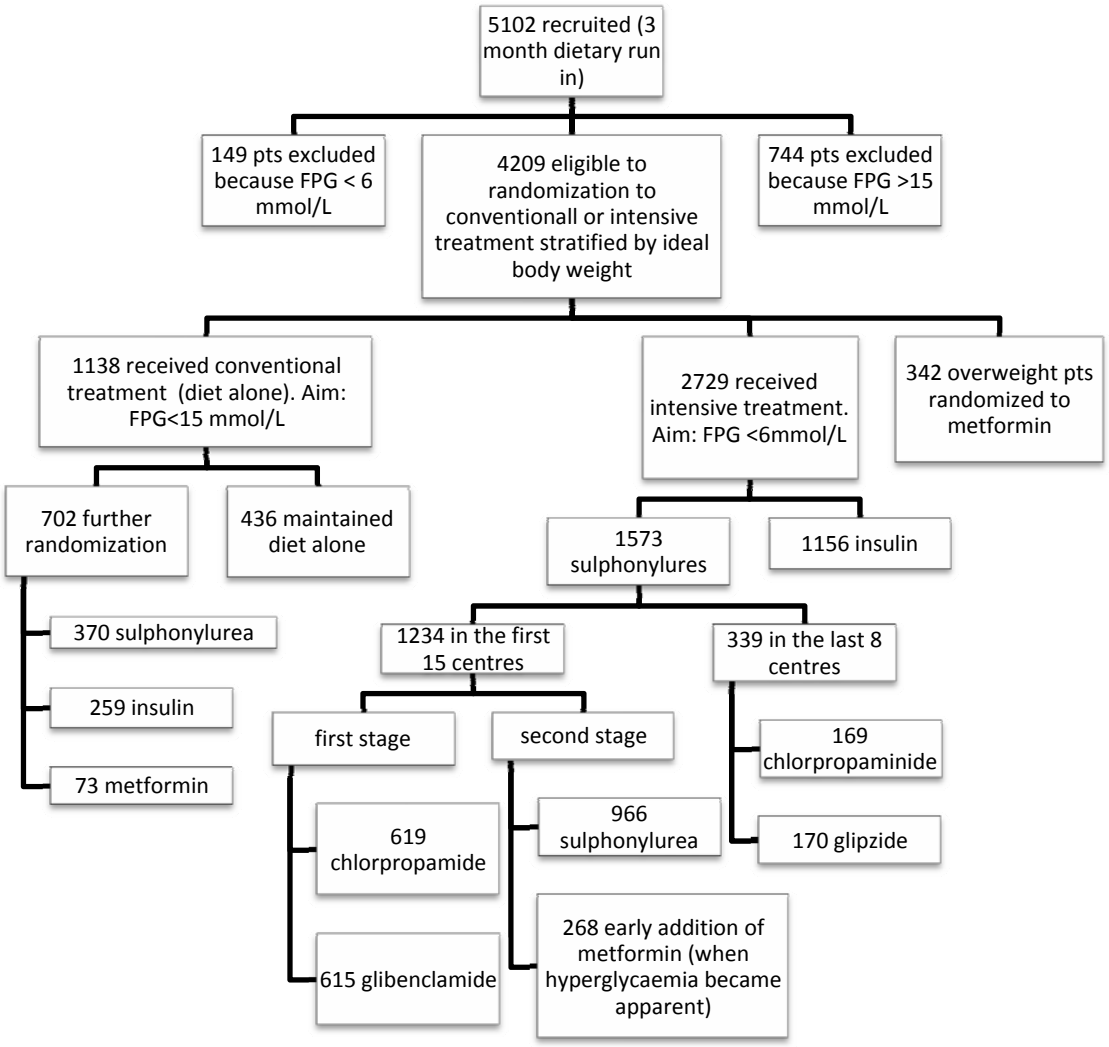
questionnaire, we assumed that his or her utilization was zero) and 12,095 quality of life questionnaires were returned (11,614 of which were fully completed).

The quality of life questionnaire was the EQ-5D, a descriptive health related quality of life instrument covering five dimensions (mobility, self care, usual activity, pain/discomfort, anxiety and depression) each of which has three levels (no problem, some problems, and extreme problems). The questionnaire also included a visual analogue scale (VAS) which prompted patients to record in a thermometer like scale between 0 and 100 their perceived overall health state.

The resource use questionnaire recorded information on all home, clinic, and telephone contacts with general practitioners, nurses, podiatrists, opticians, and dieticians, and with eye and other hospital out-patient clinics over recall periods varying from four months (Questionnaire 1, administered between 1996/1997) to one year (Questionnaires 2 to 6). The recall period for Questionnaire 7, administered at the closing of the follow up period, was patient specific and varied from one month to one year, depending on the date of completion of the subject's last questionnaire.

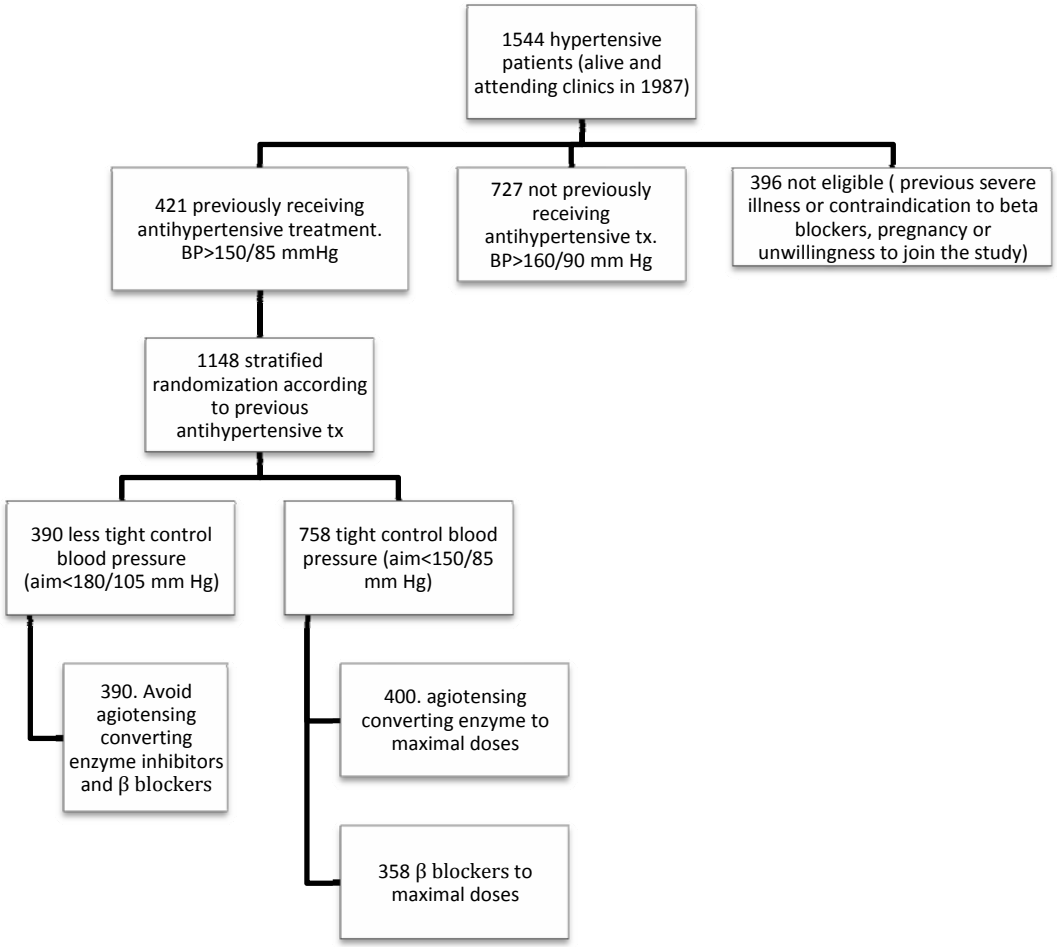
Appendix

A 1.1 Patient flow in the trial phase of UKPDS (1977-1997) – first randomization



Numbers and Structure of the trial derived from: (UKPDS(33) 1998; UKPDS(34) 1998)

A 1.2: Second randomization – Hypertension in diabetes study



Numbers and Structure of the trial derived from: (UKPDS(38) 1998)

References

Adler, A. I., I. M. Stratton, et al. (2000). "Association of systolic blood pressure with macrovascular and microvascular complications of type 2 diabetes (UKPDS 36): prospective observational study." BMJ 321(7258): 412-419.

Clarke, P., A. Gray, et al. (2002). "Estimating Utility Values for Health States of Type 2 Diabetic Patients Using the EQ-5D (UKPDS 62)." Med Decis Making 22(4): 340-349.

Clarke, P. M., A. M. Gray, et al. (2004). "A model to estimate the lifetime health outcomes of patients with Type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68)." Diabetologia 47(10): 1747-1759.

Connolly, V., N. Unwin, et al. (2000). "Diabetes prevalence and socioeconomic status: a population based study showing increased prevalence of type 2 diabetes mellitus in deprived areas." Journal of Epidemiology and Community Health 54(3): 173-177.

Ewart, R. M. (1999). "The UKPDS: what was the question?" The Lancet 353(9167): 1882.

Gray, A., P. Clarke, et al. (2002). "Implementing Intensive Control Of Blood Glucose Concentration And Blood Pressure In Type 2 Diabetes In England: Cost Analysis (UKPDS 63)." BMJ: British Medical Journal 325(7369): 860-863.

Groop, L. (2005). "Type 1 diabetes and type 2 diabetes: what do they have in common?" MEDICOGRAPHIA 27.

HMSO (1980). Classification Of Occupations. London, Office of Population Censuses and Surveys.

Holman, R. R., S. K. Paul, et al. (2008). "Long-Term Follow-up after Tight Control of Blood Pressure in Type 2 Diabetes." N Engl J Med 359(15): 1565-1576.

Lobstein, T. and R. Leach (2004). "Diabetes may be undetected in many children in the UK." BMJ 328(7450): 1261-c-1262.

Merrick D. (2005). “The PBS diabetes population prevalence model”, Heslington, UK: Yorkshire and Humber Public Health Observatory, University of York.

Morgan, C. L., P. McEwan, et al. (2006). "Characterization and comparison of health-related utility in people with diabetes with various single and multiple vascular complications." Diabetic Medicine 23(10): 1100-1105.

Nichols, G. A., C. M. Gullion, et al. (2004). "The Incidence of Congestive Heart Failure in Type 2 Diabetes." Diabetes Care 27(8): 1879-1884.

Smushkin, G. and A. Vella (2010). "What is type 2 diabetes?" Medicine 38: 597-601.

Thomason, M. J., J. P. Biddulph, et al. (2005). "Reporting of diabetes on death certificates using data from the UK Prospective Diabetes Study." Diabetic Medicine 22(8): 1031-1036.

Turner RC, Holman RR, et al. (1991). "Study design, progress and performance." Diabetologia 34(UKPDS 8): 877-890.

UKPDS(33) (1998). "Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33)." The Lancet 352(9131): 837-853.

UKPDS(34) (1998). "Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (34)." The Lancet 352(9131):854-65.

UKPDS(38) (1998). "UK Prospective Diabetes Study 38: tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes." BMJ: 703-713.

Wang, Y. C., K. McPherson, et al. (2011). "Health and economic burden of the projected obesity trends in the USA and the UK." The Lancet 378(9793): 815-825.

Wild S, Roglic G, et al. (2004). "Global prevalence of diabetes: Estimates for the year 2000 and projections for 2030." Diabetes Care 27: 1047-1053.

Chapter 2

THE IMPACT OF DIABETES RELATED COMPLICATIONS ON QUALITY OF LIFE

Abstract

This chapter reports on estimates of the quality of life impact of six diabetes related complications: myocardial infarction, ischaemic heart disease, stroke, heart failure and amputation based on EQ-5D data on 3,380 patients collected in the UKPDS Post Trial Monitoring Study over the period 1997-2007. Herein, I show that estimates based on cross-sectional data are widespread in the literature, being less expensive and easier to collect than longitudinal data, but may come at a cost in terms of biases.

The availability of seven rounds of data, over the 1997-2007 period, allows us to make more precise estimates, test whether sample selection (e.g. non response) matters, and compare cross-sectional results with those from repeated measures of quality of life over time. The variation in Quality of Life (QoL) across patients is found to be twice that observed within a patient over-time and, when this heterogeneity is correlated with the likelihood of having a diabetes-related complication, the results of the commonly used pooled estimations are shown to be biased. Cross-sectional analysis in the UKPDS sample overestimates the effects of complications on utility. Using fixed effects estimators, it is shown that variation in the quality of life between patients is strongly influenced by time-invariant patient characteristics. Also, once unobserved time-invariant heterogeneity is accounted for, selection into the sample (and hence non-response to questionnaires) ceases to be a source of bias in the sample.

Fixed effects methods however do not allow us to estimate the relation of time invariant observed patient's characteristics which may be important correlates of a patient's QoL. In this chapter therefore, panel data is linked with cross sectional data in a second-stage analysis, using patient's fixed effects as the dependent variable to examine whether certain patient characteristics, such as BMI, smoking status, exercise, alcohol consumption, gender, race, and socioeconomic status, are related to a patient's health score.

While stroke, heart failure and amputations have substantial effects on QoL, many other complications do not alter it significantly once time invariant patient heterogeneity is accounted for. The results highlight the importance of tracking QoL changes over time to distinguish between time invariant correlates of wellbeing and the effect of diabetes related complications.

2.1 Introduction

The aim of this chapter is to characterize the changes in the wellbeing of patients with type 2 diabetes over time and assess the significant determinants and predictors of this change using the EQ-5D (European Quality of Life 5 Dimensions) instrument. Reliable estimates of the quality of life impact of diabetes related complications are important for researchers conducting trial-based and model-based evaluations of the cost-effectiveness of interventions.

This chapter is divided into seven sections. Section 2.2 offers an overview of the literature aimed at comparing and providing a range of magnitudes of the effect of complications on utility tariffs found in the literature, within diabetes and other chronic conditions. It also seeks to evaluate empirical methods for health state valuations and their implications.

Section 2.3 shows how the utility measures used are derived and provides a detailed description of the data employed and the way the data set for estimation was constructed. It highlights the intrinsic relationship between the data set construct and the answerable questions, in particular, different ways of looking at a patient's history.

Section 2.4 illustrates the methodological approach to addressing selection bias in the longitudinal context. It also shows how one might construct a causal story by exploiting longitudinal information within patients.

In section 2.5 estimates of the impact of major complications on two self reported measures of Quality of Life (QoL) are presented: (1) the EQ-5D utilities derived from population based time trade-off values and (2) the visual analog scale (VAS). The long and short term effects of complications on QoL are also evaluated and the ranking of subjective (VAS) versus induced (derived index) preferences is also compared.

In section 2.6 a second stage analysis is presented in which panel data and cross sectional observations are linked in order to explain between-patient variations. Fixed effects are used as the dependent variable to see what intrinsic patient characteristics are most related to a patient's utility score.

The conclusion in section 2.7 summarises the major inferences that can be drawn from the information presented in the chapter, highlighting the limitations and contributions of the findings.

2.2 Literature review

Major biomedical electronic databases were searched in this study including *Medline* (Medical Literature Analysis and Retrieval System On-Line) and EMBASE (Excerpta Medica Database) employing key terms diabetes type 2, complications, EQ-5D and quality of life; as well as specific economic databases within JSTOR (Journal Storage) for econometric methodology. The aim of this literature review was to identify and review the econometric methods used in studies that employ the same health state evaluation as in the UKPDS. The final search was conducted in April 2011.

Methods of Health State Evaluation

In this study, the EQ-5D is used for analysis. However it should be noted that this is just one of a number of instruments that can be employed to generate preference-based values for assessing QoL. Other instruments used in the literature include Short Form (SF) 36 and SF-12, the Diabetes Quality of Life Measure and the Symptom Checklist, Health Utilities Index 3 (HUI3) and Rand-36, to name just a few. Which instrument is the most appropriate for diabetics is a matter of debate. Not all of them

can be readily transformed into a single index, and among those that can be collapsed into a single comprehensive scalar that can be compared across diseases and used for economic evaluations, the EQ-5D is the most widely used (Welch, Jacobson et al. 1997).

The choice of the most appropriate elicitation mechanism is often also controversial (Glasziou, Alexander et al. 2007). Both the standard gamble (SG) and time trade off (TTO) methods share a common theoretical foundation in utility theory, i.e., they both require people to sacrifice something they value (certainty and life expectancy, respectively) in order to gain another (quality of life). In a comparison between these two methods, Dolan et al. (1996) showed that TTO performed slightly better in terms of the internal consistency of the answers given by respondents, the sensitivity of valuations to parameters known to influence respondents, and the reliability of responses when the valuation task is repeated by the same respondents.

In the UK the most widely used weights are TTO values from a UK population study known as the Measuring and Valuing Health study (Dolan, Gudex et al. 1996), which is also the algorithm used for analysis in this chapter. Herein, respondents were asked what proportion of their remaining life they would sacrifice for being relieved of a given health state and returned to perfect health. The higher the proportion the respondent is willing to forgo, the greater the burden attributed to the health state. In the case of states worse than death, the choice was between dying immediately and spending a length of time *x* in the target state followed by 10-*x* years in the "full health" state. The more time required in "full health" state to compensate for a shorter time in the target state, the worse the target state. One caveat to this study is that the 10 year horizon chosen by the authors may not be long enough for young respondents to make meaningful trade-offs. At the same time, older respondents in the study had much lower

valuations which may be an artefact of the TTO method as well. For states that were rated as better than death, if older patients did not believe they had 10 years of life expectancy, they might willingly give up these 'excess' life years, thereby depressing the apparent value attached to the health states. The authors argue however that using variable time horizons based on a person's own expected life would create even greater problems of measurement and interpretation. Thus, one of the most interesting findings is the effect the age of the respondent had on health state valuations. The valuations of 'middle-aged' respondents appear to be higher than those of younger respondents, whilst older respondents have much lower valuations than those in the other two groups. This may lend support to the notion that the middle-aged have the lowest rates of time preference and thus place relatively more weight on years in the future (i.e., the ones they are being asked to sacrifice) than either younger or older respondents. However, the fact that the effect of age is not uniform across all states, being more pronounced for moderate and severe states than for mild ones, suggests that the effect of time preference for health is not independent of the severity of the health state.

There is no consensus on whether individuals unaware of the disease under study or patients suffering from the disease provide more appropriate weighting for utility data (Wit, Busschbach et al. 2000). Valuations used in decision analysis and cost-utility analyses can be given by different groups, such a patients or members of the general public. In the UK, reimbursement agencies such as the National Institute for Health and Care Excellence (NICE) recommend utility data obtained from societal preferences (NICE 2004). Among meta-analyses that investigate the effect of response groups using the EQ-5D, Dolders et al. (2006) found no difference in valuations between

members of the general public and patients. In contrast, Peeters and Stiggelbout (2010) report a difference in utility between the group means equal to Coehen's d=0.18 (SD: 0.36), meaning that patients have higher valuations than non-patients and that the difference is larger among studies using TTO than studies using SG.

Health States in diabetic patients

This chapter builds on previous work by Clarke, Gray et al. (2002) who employed a single cross-sectional EQ-5D survey from the same study using 3,192 UKPDS patients after 10.6 mean years follow up to estimate the short- and long-term effects on QoL of six pre-specified clinical events: myocardial infarction (MI), ischaemic heart disease (IHD), stroke, heart failure, amputation, and blindness. There are now a substantial number of studies reporting utility scores for people with diabetes and for a number of complications associated with the disease, but with the exception of the the Fenofibrate Intervention and Event Lowering in Diabetes (FIELD) Study, the action in diabetes and vascular *disease:* preterax and diamicron MR controlled (ADVANCE) evaluation, the cost of diabetes in Europe type 2 (CODE-2) study and UKPDS, the majority of diabetes trials available lack a direct, single measure of wellbeing[1].

The literature reporting quality of life among patients with diabetes could be divided into studies looking at the relation between therapies and quality of life and, those looking at how the occurrence of complications affects quality of life. A more holistic approach, given that complications could never be experimentally assigned in the fashion of a randomized controlled trial (RCT), should include all potential drivers of

[1] There are a number of quality of life instruments with associated mapping and reference algorithms. The EQ-5D however is perceived as being the preferred method, not only due to its simplicity and ease of administration but also because it produces greater mean differences, specially for macro vascular complications while also showing a lower signal to noise ratio compared to measures derived from mapping (Glasziou, P., J. Alexander, et al. (2007)).

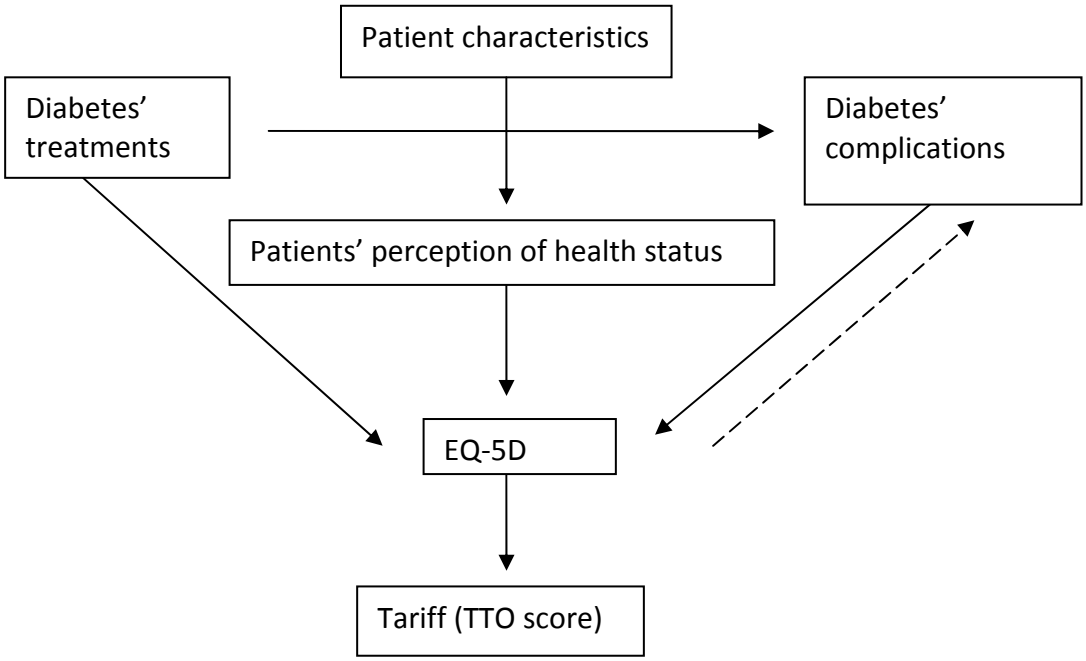
utility as illustrated in Figure 2.1; specifically the impact of therapies on the likelihood of complications and the impact of complications on quality of life as well as the direct impact of treatments themselves on quality of life, assuming treatments affect quality of life through side-effects. No study to my knowledge takes this structural approach, and I too attempt a reduced form specification looking only at the right hand side of the Directed Acyclic Graph (DAG) (Figure 2.1). The reason for this is twofold: 1) after 1997 we no longer know the treatments followed by patients; all we know is that they follow best standard practice. The underlying assumption is that new protocol for the management of the disease stemming from the UKPDS 33 and UKPDS 35 publications was already in place by the time patients entered the follow up phase. 2) Although UKPDS 33 has shown that patients allocated to both intensive glucose and tight BP control regimens had fewer events than those allocated to either regimen alone or neither, none of the intensive anti-diabetic agents showed an advantage over the others (UKPDS 37).

Because in the UKPDS no significant relation between treatments and quality of life was found, any possible existing link between treatment and the likelihood of complications will be part (as are many other things out of the researcher's control) of the idiosyncratic error term.

The dashed line in Figure 2.1 illustrates the possibility of reverse causality in the modelling of quality-of-life impacts of diabetes-related complications. Clarke and colleagues (2009) show that index scores based on the EQ-5D are a predictor of vascular events and other major complications, including mortality in people with type 2 diabetes in the FIELD study. The incidence rate of some major complications and the risk of mortality for a patient with an EQ-5D index score of 0.60 is almost double the rate of

those with an index score of 1.00, after adjusting for other known determinants of event risk.

Figure 2.1: Directed Acyclic Graph of factors believed to affect health related quality of life



The evidence for the impact of treatments on QoL is mixed. Mayou et al. (1990), UKPDS 37 (1999), and Weinberger et al.(1995) report no significant relation between therapies that control blood glucose and quality of life. Tabaei et al. (2004) and Lundkvist (2002) show instead a significant relation between therapies and quality of life. Mayou et al. (1990) studied a group of 121 non insulin dependent patients, recruited by the UKPDS Oxford centres and randomized to diet, tables or ultralente insulin therapy. In addition, they also recruited 57 insulin dependent patients. Both groups were administered a social difficulty questionnaire covering aspects of leisure, work and family life as well as a mood questionnaire (POMS) focusing on tension, depression, fatigue, anger and lack of vigour. The type of therapy made little difference to phychological, social or attitudinal variables. In the UKPDS 37 study the mood

disturbance (Profile of Mood State), cognitive mistakes (Cognitive Failures Questionnaire), symptoms, and work satisfaction questionnaire as well as a generic questionnaire (EQ5D) were administered. Allocated therapies were neutral in effect, with neither improvement nor deterioration in QoL scores for mood, cognitive mistakes, symptoms, work satisfaction, or general health. Weinberger et al (1995) differed from the above studies in that the authors did not look at the effect of drugs but at the effects of adjunct care on quality of life. Specifically they instituted a minimum of one monthly nurse-initiated telephone contact to provide patient education (with special emphasis on regimens and significant signs and symptoms of hyperglycemia and hypoglycemia) in order to reinforce compliance with regimens, monitor patients' health status, facilitate resolution of identified problems, and facilitate access to primary care. Although the intervention modestly improved glycemic control, it did not have an effect on health related QoL (SF-36) or diabetes-related symptoms.

Both Tabaei et al. (2004) and Lundkvist et al. (2005) found that after controlling for age, gender, and complications, frequency of hyperglycemic symptoms often associated with insulin's intake was negatively associated with derived health utility scores in type 1 and type 2 diabetes.

A comprehensive systematic review of 45 published studies on the effects of diabetes (type 1 and 2) and the effects of complications on quality of life exists in Lung et al. (2011). Table A2.1, in the appendix, summarizes their findings in terms of number of studies per complication, mean utility score, range and mean number of subjects. Thirty three of these studies used preference-based measures. Lung et al. find substantial heterogeneity in utility values across studies but point out the irrelevance of the differences for most practical purposes. For instance by using simulated

hypothetical therapies the authors find that different scores had little impact on estimated quality-adjusted life years (QALYs).

Among studies looking at how the occurrence of complications affects quality of life, I have restricted attention to studies that use the EQ-5D in patients with type 2 diabetes. These are summarized in Table 2.1 according to ascending year of publication. All the studies have collected one, maximum two (non-overlapping) rounds of data. Blindness and amputation are associated with the lowest utility values, on average, 0.53 and 0.56 respectively.

The most common methodological approach was to obtain differences in means between samples with particular characteristics or by treatment allocation. The most widespread form of multivariate analysis was Ordinary Least Squares (OLS) which is essentially a controlled difference in means when the covariates of interest are binary.

In seeking to identify correlations, the primary aim of the studies available becomes one of incidence rather than causation. If the aim is to be able to eradicate complications, not being able to construct a causal argument in this area is limiting because the inferences that can be made on the impact of complications may not hold across subjective groupings. For instance QoL may be higher among men than women, with women having more complications than men because of greater longevity. This same QoL could vary within gender according to race, with Caucasians having fewer complications than Asians but Asians reporting higher QoL as a cultural phenomenon, or with socioeconomic status, with higher earners having fewer complications and higher QoL. If these potential correlations across covariates are not accounted for, or when some of the grouping affecting utility and the likelihood of complications are not

observable, one cannot make universal claims on the actual impact of complication on QoL.

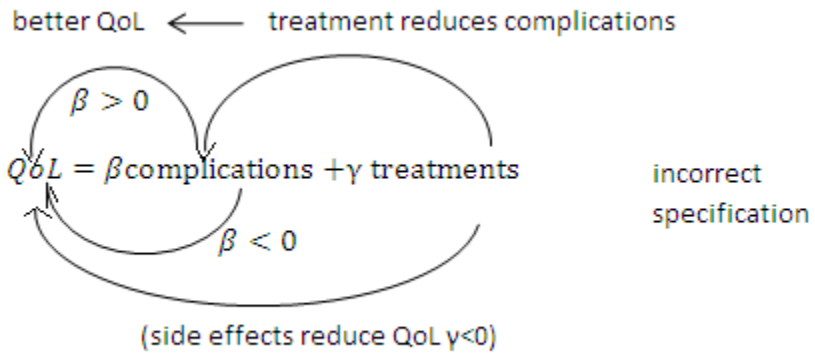
The disutility decrement for stroke in published studies, for example, varies from 0.164 (Morgan, McEwan et al. 2006) to 0.04 (Glasziou, Alexander et al. 2007). It is important to ask whether this is a clinically significant difference. Walters et al.(2005) found that the minimal important difference (MID) for the EQ-5D in patients with chronic diseases was between 0.011 and 0.014.

The use of cross-sectional data is widespread as it is less expensive and less difficult to collect than longitudinal data. However, a single observation per patient is not ideal for estimating the impact of long-term complications. Obtaining evidence for robust ranking of different complications and their severity requires large observational studies extending several years. From the viewpoint of a single questionnaire complications may weight more heavily healthy individuals that have survived complications, particularly if they are more likely to participate than the rest of their cohort, leaving selection bias unaccounted for. Moreover, cross-sectional differences between average measures of utility for different health states are variations across the population at a point in time rather than changes in an individual's health related QoL (HRQoL) were the individuals to experience a transition between these states. For example, if the average utility for diabetic patients with no history of complications is 0.72 and that with complications is 0.62, the difference of 0.1 between these levels is often regarded as the impact of complications on QoL. Ideally, measures of *change* should be based on the average difference in measures of QoL at two or more points in time that are collected for the same patients before and after a complication occurs. In essence, measures of change should stem from longitudinal data.

More than half of the studies cited herein (Table 2.1) use stepwise inclusion criteria to select coefficients. Not only are there better criteria for the inclusion of model covariates (AIC, R-squared, estimation theory) but also by adding one variable at a time it is possible to not only miss the optimal model but also to lower the P-value of the remaining predictors by removing variables that do not make the arbitrary threshold. This approach fundamentally highlights the lack of theory behind what may be the factors influencing quality of life.

One study among those cited included both complications and treatments using a multivariate approach (Bagust et al. 2005), which poses specification issues. The authors find that treatments are negative and statistically significant predictors of QoL, in magnitude bigger (insulin: -0.049, $P< 0.011$) than coronary heart disease (CHD), stroke, neuropathy and retinopathy. Figure 2.2 shows why including treatments and complications in the same equation may bias the coefficient for treatments (if treatments have an effect on the likelihood of complications). The management of diabetes is geared towards reducing the likelihood of complications through a combination of life-style improvements and medication. If treatments also increase QoL through avoiding complications, while having side effects decreases QoL, then including both in the regression will only capture the impact of the side effects as the causal path from complications to QoL is blocked. Adding interaction effects between these two variables will not solve the problem either. If one cares about the impact of treatments on QoL, there is no need to include complications; a reduced form specification will suffice. On the other hand, if a fully structural approach is sought, a separate equation for the impact of treatment on complications should be included in the model.

Figure 2.2: Misspecification issues arising from including both treatments and complications in a single equation



Methodological issues aside, very few analyses give separate coefficients for multiple complications. However, Clarke (2002), Bagust (2005), Morgan (2006) and O'Reilly (2010) do this. The remaining studies focus on eye related complications or cardiac complications only. In Ose et al. (2009) complications are grouped together even though these are as diverse as coronary heart disease, cancer, heart failure, heart attack and osteoarthritis. This may not be surprising since the incidence rate of events is between 5%-15% in the population, which in small samples may not be enough to discriminate between all possible and different complications.

Table 2.1: Overview of studies published up to 2010 on the effects of complications on QoL using the EQ-5D

Author(s), publication year	Country of origin and trial	Elicitation Mechanism (TTO, SG) for EQ-5D index and other instruments	Number of subjects and questionnaire rounds administered	Econometric Method	Effect
Clarke et al. (2002)	United Kingdon. UKPDS	EQ-5D using TTO UK weights.	3,192 patients. 1 round	Tobit and CLAD	Myocardial infarction=–0.055 Blindnessin1eye=–0.074 Ischaemic heart disease=–0.090 Heart failure= –0.108 Sstroke=–0.164 Amputation= –0.280 Baseline (no complications)=0.814 (all significant at the 5% level)
Redekop et al. (2002)	Netherlands. CODE-2 study	EQ-5D using UK TTO weights.	1,224 patients, recruited from 29 GPs	Ordinary Least Squares (OLS)	Mean tariff=0.75. OLS intercept=0.986. Micro vascular=-0.068. Macro vascular=-0.077. Both micro and macro=-0.169
Lundkvist et al. (2005)	Sweden. Patients under insulin and/or oral medication in 7 centres.	EQ-5D using UK TTO weights.	309 patients. 1 round	Difference in means	Tariff with hypoglycaemia= 0.70 (SD=0.27). No hypoglycaemia=0.77 (SD=0.24)
Bagust et al. (2005)	Belgium, Spain, Italy, Netherlands and Sweden. CODE-2 study	EQ-5D using UK weights.	4,641 patients. 1 round	OLS	Similar results for OLS and CNLM. Constant (outside range)=1.027 CHD=-0.028 Stroke=-0.115 Proteinuria=-0.048 End stage renal disease=-0.175 Neuropathy=-0.084 Peripheral vascular disease=-0.06 Foot ulcer=-0.170 Amputation=-0.272 Retinopathy=-0.057 blindness=-0.061
Bharmal et al. (2006)	U.S. Medical Expenditure Panel Survey	EQ-5D , SF-6D, PCS-12, MCS-12 and the EQ-VAS were compared to assess ceiling effects.	11,248 patients, 165 diabetics with no medical conditions. 2 non-overlapping rounds	Cross sectional, changes in mean values	EQ-VAS: 79.59 (SD=1.74) when EQ-5D=11111. Mean difference in tariff between diabetics and non-diabetics: 0.51

Morgan et al (2006)	Cardiff and the Vale of Glamorgan. Health Outcomes Data Repository.	EQ-5D using UK weights.	4,502 patients answered mailed interviews six weeks after hospital discharge (33% response rate). For patients with multiple questionnaire returns, the most recent questionnaire returned was used.	GLM (OLS)	Constant (outside range)=1.608. Coronary heart disease:-0.066 Stroke=-0.114 Diabetic foot=-0.069 ESRD=-0.082 PVD=-=0.063 Retinopathy=-0.029
Sakamaki et al. (2006)	Japan. Survey at a hospital in the Saitama Prefecture.	EQ-5D	220 hospital patients. 1 round	ANCOVA	Overall mean=0.86. No significant relation between EQ-5D scores and the presence of complications. Neuropathy=0.85 (without=0.87) Retinopathy=0.86 (without 0.87) Nephropathy=0.81 (without=0.87)
Farmer et al. (2007)	Oxfordshire and South Yorkshire. DiGEM (The diabetes glycaemic education and monitoring study)	3 arms: standard care with measurements of HbA1c every 3 months; blood glucose self-monitoring with facultative feedback and self-monitoring with training	453 patients with non-insulin treated type2 diabetes. 1 round collected at baseline and 12 months follow up.	Difference in means	No evidence of a significantly different impact of tx on the different sub-groups in terms of EQ-5D. Average score across the sample: 0.814
Matza et al. (2007)	England, Scotland. Surveys.	Standard Gamble (raw scores), EQ-5D, PGWB, and Appraisal of Diabetes Symptoms disutility of diabetes medication-related attributes (weight change, gastrointestinal side effects, fear of hypoglycemia)	129 patients, 1 round	Correlation between SG rating and EQ-5D, PGWB and ADS were performed to assess the validity of SG procedure.	Average utility (EQ-5D)=0.76. Utility of diabetes without Complications=0.89. Correlation with SG =0.31 p<0.001 11 pair-wise comparisons of health states with different levels and compositions of side effects, all compared to the patient's baseline health state.

Sakthong et al. (2008)	Thailand. Survey of T2D in a Bankok hospital	Comparison of EQ-5D index scores using the UK, US, and Japan preference weights in a Thai sample	303. one round and 64 patients re-sampled at random for test retest	Difference in means	HbA1c<7 : UK=0.79, US=0.83, Japan=0.77 HbA1c≧7: UK=0.75, US=0.80, Japan=0.73 Neuropathy (no): UK=0.81, US=0.84, Japan=0.79 Neuropathy (yes): UK=0.69, US=0.77, Japan=0.69 Netinopathy (no): UK=0.78, US=0.82, Japan=0.75 Netinopathy (yes): UK=0.67, US=0.75, Japan=0.68 Nephropathy (no): UK=0.77, US=0.82, Japan=0.75 Nephropathy (yes): UK=0.67, US=0.75, Japan=0.68 Cardiovascular disease (no): UK=0.78, US=0.88, Japan=0.76 Cardiovascular (yes): UK=0.65, US=0.73, Japan=0.66
Sullivan, Ghushchyan et al. (2008)	USA. Medical Expenditure Panel Survey (MEPS)	SF-12's PCS-12 and MCS-12. EQ-5D and VAS. The scoring algorithm for the EQ-5D uses U.S. Community based preferences.	43,221 patients recruited through MEPS, 6.3% with diabetes.	CLAD	normal weight with DM=0.761 EQ-5D index. DM obese=-0.057. Hyperlipidemia normal weight=-0.007. Hyperlipidemia obsese=-0.050. Hypertension normal weight=-0.016. Hypertension obese=-0.055
Ose et al. (2009)	Germany. Disease management program (DMP) within Evaluation of a Large-Scale Implementation of Disease (ELSID) study (2005–2007)	Compare the effectiveness of care provided by the DMP with that of routine care using tariffs from Greiner (2003) 6 country study.	3,546 patients (42.2% response rate). 1 round	ANCOVA	Tariffs with T2D, no complication under routine care: 0.79(SE=0.036). DMP: 0.83 (SE=0.027). Differences is means as presented for patients with 1, 2 and up to >6 complications
O' Reilly (2010)	Ontario, Canada. Survey.	U.S. scoring algorithm for the EQ-5D.	1,143 patients answered to the mailed questionnaire. Complications were self reported.	OLS	MI=-0.050 (P<0.005) Amputation=-0.063 Stroke=-0.046 (P<0.005) Kidney failure=-0.102 (P<0.005)

General Approaches to Modelling Utility

Health utility data often show an apparent ceiling effect, where a proportion of individuals achieve the upper value of 1. There are, however, a number of confusing statements regarding terminology. Pullenayegum et al. (2010), for example, refer to this phenomenon interchangeably as "censoring," "truncation," or a "ceiling effect." These however are quite distinct issues and how we choose to interpret them has modelling implications. Censoring occurs when data on the dependent variable are lost or limited *but* no information on the regressors is missing. Truncation occurs when some observations on *both* the dependent variable and regressors are lost. Consequently, truncation entails a greater loss of information than censoring (Wooldridge 2002).

The term ceiling effect itself has two distinct meanings, referring to the level at which an independent variable no longer has an effect on a dependent variable, or to the level above which variance in an independent variable is no longer measured or estimated. Although the latter part of that definition is not entirely appropriate in this case, I shall refer to the probability mass accumulating at the upper limit as a ceiling effect. This is chosen out of the belief that the issue does not revolve around data observability, but rather because we are interested in features of the distribution of the tariff given a set of covariates, such as $E(tariff|X)$ and $P(tariff = 1|X)$.

It has been argued that the non-trivial probability mass accumulating at the upper limit of the outcome measure renders Ordinary Least Squares (OLS) inappropriate or of limited value. OLS implies constant partial effects so that the probability of a tariff value exactly equal to one is very small relative to the observed distribution. Another drawback is that it does not guarantee that the predicted value will fall within the observed boundaries (i.e., the predicted tariff values may be greater than 1). This is also

a problem with feasible generalized least squares (FGLS) procedures such as Random Effects (RE) or within regression estimators, like Fixed Effects (FE). This should not be a concern when the estimation features predominantly dichotomous explanatory variables as in our case (i.e., has the patient had an event? yes or no).

As an alternative to OLS and FGLS, different approaches have been suggested in areas other than diabetes for data with similar characteristics. Examples include Tobit (Clarke, Gray et al. 2002; Sullivan and Ghushchyan 2006); Two Part Models (Mullahy 1998); Latent Class estimators (Pullenayegum, Tarride et al. 2010); and Censored Least Absolute Deviations (CLAD)(Powell 1984).

CLAD is known to be robust to heteroskedasticity and non-normality. In the literature geared to inform economic evaluations, however, the aim is to provide estimates of the mean. Cost effectiveness models use the overall aggregated effect associated with a particular event across the whole population, regardless of the degree of skewness. On these grounds, the CLAD regressor may not be appropriate as it gains the desirable property of consistency through the use of median regression and as such does not account for the strength of all individuals.

One should not necessarily believe that the "true" measurement exceeds 1 in order to apply a Tobit model. Amemiya's 1985 taxonomy refers to the modelling of the distribution of observed values as type I Tobit models. Wooldridge (2002) calls the outcome of such models "corner solutions", meaning an intended optimal choice that happens to be at either end of the distribution. Studies that shy away from Tobit because they do not want to consider "supra-normal" states should be aware that Tobit is also applicable where one wishes to make a more realistic interpretation of the instrument and to recognize that the tariff is just not able to capture small departures

from perfect health. Tariffs have gaps by construction at the upper as well as the lower level but at the lower limit the effect is less important as the number of observations tends to be very small. As soon as a response departs from perfect health (equivalent to a score equal to 1 in all five dimensions of the EQ-5D or 11111), using the Dolan's algorithm, we need to subtract first the constant term (-0.081) and the subsequent lowest possible deduction is 0.036 if the person reports a 2 in usual activities (and thus 11211), which means that the interval 1 – 0.883 is empty (0.081+0.036=0.117).

The Tobit model makes the same assumptions about error distributions as the OLS model, but is much more vulnerable to violations of those assumptions.
In an OLS model with heteroskedastic errors, the estimated standard errors can be too small. If the disturbance is either heteroskedastic or non-normal, then the Tobit estimates would be inconsistent. It is possible to get consistent estimates with heteroskedastic errors if the heteroskedasticity is modelled directly. But this is hard to do accurately. In Tobit, the conditional mean of the dependent variable $E[y|y < \tau]$ is a non-linear function of the truncation point (τ), the variance (σ), the covariates(X) as well as the parameters (β). Because it depends on the variance, with heteroskedasticity, the conditional mean is no longer unique.

$$E[y|y < \tau] = X\beta + \sigma \left[\frac{\phi\left[\frac{X\beta}{\sigma} - \tau\right]}{1 - \Phi\left[\frac{X\beta}{\sigma} - \tau\right]} \right]$$

where ϕ is the probability density function and Φ denotes the cumulative density function.

As Sigelman and Zeng (1999) point out, however, there are circumstances in which one may not be interested in the observed conditional mean but on the latent or unconditional dependent variable. If one chooses to focus on the latent variable

$E[y^*] = X_i\,\beta$ heteroskedasticity does not lead to bias. This approach is not particularly useful with tariffs because the variable of interest is the observed tariff and not the hypothetical latent tariff. Utility cannot exceed 1 in this context because one cannot do better than full health.

Two part models have also been used in the literature. The idea there is to model the sample into two separate components, one with higher utilities and the other with lower utilities, so that the distribution of tariffs in the total sample is a mixture distribution:

$$\mathrm{E}(tariff_{it}|x_{it}) = \Pr(tariff = 1) \times 1 + (1 - \Pr(tariff = 1)) \times E[tariff_{it}|tariff_{it} < 1, x_{it}]$$

The first part (selection equation) describes the probability of being in full health:

$P(full\ health_{it} = 1|X_{it}) = \frac{\exp(\alpha X_{it})}{1+\exp(\alpha X_{it})}$, where α is the vector of estimated coefficients.

The second part (substantive equation) is a model of the expectation of the health state utility level conditional on it being in a less than perfect health state:

$E(tariff_{it}|tariff_{it} < 1, X_{it})$.

If any factors that affect both the probability of reporting full health and the outcome of interest are omitted from the model, these will enter the error term in both equations and induce correlation among them. If omitted variables are likely to be a problem then the estimates will be biased, with the size of the bias depending on the correlation between the errors in the two models. Unfortunately, the independence of errors cannot be tested directly.

Other potential options that could be adopted to the characteristics of this data are Fractional Logit (Papke and Wooldridge 2008) and Pantob (Honore 1992) .The Fractional Logit model (Flogit) was first proposed by Papke and Wooldridge (1996) as an

alternative to OLS and Tobit. Flogit combines features of logit transformation and binomial regression approaches, i.e., it considers the dependent variable a count (with strictly positive values) rather than a continuous dependent variable. Levy et al. (2008), for example, use Flogit to measure the impact on utility for different frequencies of hypoglycaemia episodes (their elicitation mechanism never leads to negative utilities). Because we have tariff<0, in order to apply Flogit, we would need to work with the transformed variables: $q = \frac{tariff - a}{1 - a}$, where $a < 0$ represents the lower limit, so that $q \epsilon [0,1]$. Predicted values for q can then be retransformed to produce predicted values for the underlying untransformed variable tariff=q*(1-*a*) +*a.* For a panel specification of Flogit one could follow the Papke and Wooldrige (2008) use of the generalized estimating equation (GEE) approach. The Flogit model ensures consistency in prediction (i.e., all fractions sum to one). Also, the logit transformation allows a diminishing marginal effect of any explanatory variable on the conditional expectation of q_{it} due to the fact that the first derivative of $E(q_{it} \mid x_{it})$ with respect to x_{it} is close to zero as x_{it} approaches infinity.

Pantob is a generalization of Powell's (1984) proposed semi-parametric estimator for Tobit models that exploits the symmetry in the distribution of the dependent variable. By arbitrarily censoring (trimming) the upper tail of the distribution of the dependent variable (the part symmetric to the lower part of the distribution which is unobserved), the remaining observations will be symmetric. The parameters can be found by paralleling those of the OLS estimation. Honore (1992) generalized this idea for panel data allowing for fixed effects. As a feature of this semiparametric estimator, he established the consistency and asymptotical normality without maintaining parametric assumptions on the error terms. This panel Tobit estimator

(Pantob) has been used by Grootendorst (1997) to investigate the effect of insurance on drug prescriptions. The main advantages of using Pantob, as suggested in Grootendorst's paper, are that the estimator requires far less structure on the form of the distribution of disturbances than the maximum likelihood estimates. This feature is attractive because the Tobit model is sensitive to the assumptions of normality and homoskedasticity across both individuals and time periods. The estimator uses an iterative linear programming algorithm (ILPA). With ILPA, convergence occurs when there are no predicted values exceeding the boundaries of the upper and lower censoring value in two consecutive iterations.

It is not straightforward to come up with an unambiguous criterion to identify the correct statistical model where the data generating process is unknown. Two considerations shall be borne in mind while choosing the appropriate regression model: in this study: lack of bias and efficiency. In particular, Table 2.2 examines the merits of different statistical models in dealing with three problems specific to quality-of-life data: time invariant unobserved patient characteristics correlated with the presence of complications, heteroskedasticity, and ceiling effects. As each estimator carries a set of assumptions, choosing one with less restrictive assumptions represents an additional desirable characteristic.

Table 2.2: Methodological approaches to utility data

Method	Heteroskedasticity $Var(e_{it})=E(e_{it}^2)\neq\sigma^2$	$E(u_i\ x)\neq 0$	Probability mass accumulating in the upper limit	Other issues brought by the method: additional restrictions
Pooled OLS with clustered errors	✔ Standard errors can be adjusted to be asymptotically robust to both heteroskedasticity and serial correlation. Alternatively feasible generalized least squares (FGLS) can be used	✘ does not account for time invariant unobservable typical to the panel setting	✘ not well suited to impose a bounded effect of the covariates on tariffs.	
Panel FE, RE	✔ FGLS offers more elegant ways to account not only for heteroskedasticity but also serial correlation.	✘ ✔ RE accounts for this partially while FE deals with the porblem fully. However in the case of singletons, with FE one de facto discards all those observations.	✘ same as above	
Tobit, Panel Tobit RE	✘ Tobit requires normality of the error term and constant variance. But it is possible to calculate robust SE. Clustering on the panel variable produces a consistent VCE estimator when the disturbances are not identically distributed over the panels or there is serial correlation, but relies heavily on homoskedasticity.	✘ Assumes strict exogeneity and if this is not satisfied becomes an inconsistent estimator like pooled OLS. In the case of RE the fraction by which the bias is carried forward is reduced by: $1-\sqrt{\frac{\sigma_e^2}{\sigma_e^2+\sigma_u^2}}$	✔ Tobit assumes that the observed utilities are censored at 1 and hence that the true utility can be greater than 1. One can correct for this easily by calculating marginal effects within the observed boundaries.	
Wooldridge's 2005 Tobit FE	✘ same as above, however controlling for the time invariant error term may reduce the source of bias.	✘ ✔ the FE is not obtained through differencing but thorugh controlling for demeaned time varying covariates.	✔	
Fractional Regression –Flogit (Papke and Wooldridge 1996)	✘ ✔ the errors are allowed to have bounded heteroskedastic structures.	✘ no within estimator or first difference is possible when the estimator is non-linear.	✔ while holding the theoretically advantageous property of bounding utility at 1, the coefficients are less straightforward to interpret. It requires a transformation of the dependent variable in the 0-1 interval.	A binomial stochastic structure results in too much weight being given to the tails relative to the truth. So in some cases, standard errors may be too wide.
Panel Flogit (Papke and Wooldridge 2008)	✘ ✔ same as above	✘ same as above	✔ same as above	same as above
Two part model (logit + glm) cross section	✘ like any non-linear estimator, it is biased when heteroskedasticity is present	✘ same as above	✔	Requires that the error terms in the selection and outcome equations are uncorrelated.
Two part model (xtlogit + GEE) panel	✘	✘ ✔	✔	Same as above

2.3 Data

The EQ-5D questionnaire

In the UKPDS, quality of life was measured using the EQ-5D and a visual analog scale (VAS)[2]. The former is a descriptive system covering 5 dimensions: mobility, self care, usual activity, pain/discomfort, anxiety and depression, each of which has 3 levels: no problem, some problems, and extreme problems. The latter provides important information on the individual's subjective perception of their generic health status. Each composite health state in the EQ-5D has a five digit code number corresponding to the relevant level of each dimension, with the dimensions always listed in the same order. Thus a patient self reporting 11223 in the EQ-5D questionnaire means that the subject has no problems in walking about, no problems with self-care, some problems with performing usual activities, moderate pain or discomfort and is extremely anxious or depressed. Reference values for each of the possible 243 health states (3^5) are combined into a single utility score with weightings derived from a UK population study (Dolan 1997) where preference values are elicited using the time-trade-off method. The algorithm (Appendix A2.3) produces tariffs ranging from 1, for full health, to -0.594 for severe problem. The minimum value, -0.594, corresponds to the worst possible health state (i.e. 33333). As zero represents death, negative values of the index indicate states that are deemed to be "worse than death".

[2] For a copy of the questionnaire see Appendix A2.2.

The index that can be derived from the EQ-5D facilitates the assessment of a variety of health interventions through quality adjusted life years and thus cost effectiveness analysis. Hereafter, I shall refer to it as a tariff or utility, interchangeably.

A maximum of seven questionnaires are available. With the exception of the first questionnaire (administered between October 1996 and September 1997) and the last (administered in October 2007 to all surviving participants), questionnaires sent between 2003 and 2007 were administered on an annual basis coinciding with the anniversary of the study therapy decision for each patient.

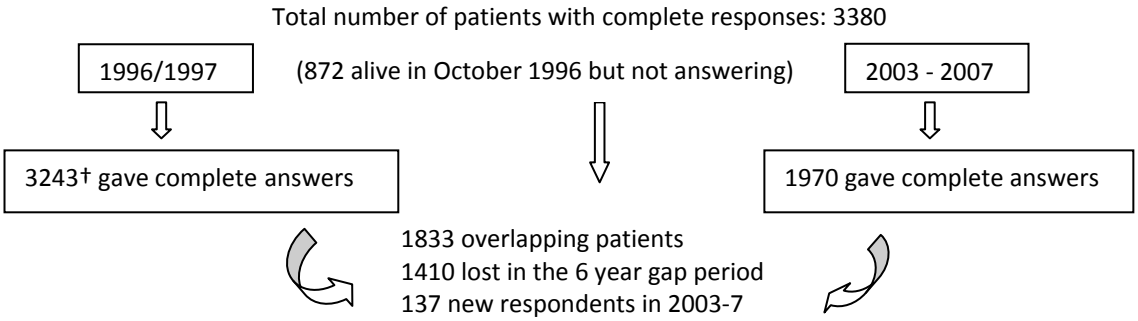
Patients were asked to complete the EQ-5D questionnaire during routine visits to UKPDS clinics, in a quiet room with no help from nursing staff, family, or friends. Patients who did not attend clinics during the survey period were sent the questionnaire by mail, and those who did not return a given questionnaire were sent up to two reminders at each round of data collection (Clarke, Gray et al. 2002).

A total of 3,380 out of the 4,252 UKPDS patients alive during the follow up period returned at least one complete EQ-5D questionnaire (Figure 2.3 illustrates the patients' flow across questionnaires). The response rate for the first questionnaire was 90% (Clarke, Gray et al. 2002). The response rates for the questionnaires administered during the 2003-2007 period were slightly lower (74%), as shown in Appendix A2. 4.

Table 2.3 shows the 25 most common patterns of response among patients that submitted at least one complete questionnaire, where 1 indicates that the questionnaire has been fully completed and a dot indicates that the questionnaire was either unreturned or incomplete. Approximately 42% of respondents answered the

questionnaire only once. The average number of questionnaires completed per patient was three.

Figure 2.3: Patients' flow across questionnaires with complete 5 level EQ-5D responses



† The UKPDS 62 reports tariffs for only 3192 patients as this is the number of patients reporting both the 5 dimensions *and* the VAS.

Table 2.3: Pattern of responses among patients who returned at least one complete questionnaire

Freq.	Percent	Pattern
1410	41.72	1.....
586	17.34	1111111
216	6.39	11111.1
101	2.99	11.....
89	2.63	11111..
81	2.4	111....
73	2.16	1111...
64	1.89	111111
61	1.8	1.11111
43	1.27	..111111
39	1.15	1111.11
38	1.12	1.1....
32	0.95	111.111
26	0.77	1..1..
23	0.68	1111..1
22	0.65	1.11...
21	0.62	11.1...
21	0.62	11.1111
19	0.56	111.1.1
17	0.5	1.111.1
15	0.44	1..1111
15	0.44	1.1111.
14	0.41	1.111..
13	0.38	..11111
13	0.38	1.....1
328	9.7	(other patterns)
Total patients:		
3380	100	XXXXXXX

The proportion of patients reporting no problems fell markedly between the first and second questionnaire. Changes across questionnaires two to seven were less pronounced. Questionnaire six, as shown in Table 2.4 has, by construction, a lower number of reported answers (1,041) than questionnaire seven (1,262) because patients with anniversaries of therapy decision in October, November and December were not mailed questionnaire six, whereas all patients alive and consenting in October 2007, regardless of their date of therapy decision, were sent a questionnaire. The last questionnaire was to act as the final state for that subject. Thus there is a slight underweighting of eligible patients with anniversaries in the later months since the process of allocating therapy decision dates was random. For all intents and purposes this should be of no concern.

Table 2.4: Proportion of levels 1, 2 and 3 by EQ-5D dimension and questionnaire round

EQ-5D DIMENSION	level	Questionnaires							
		1	2	3	4	5	6	7	TOTAL
	1	63.6%	49.9%	47.2%	45.0%	42.3%	41.4%	42.6%	50.3%
	2	36.1%	49.7%	52.2%	54.4%	57.0%	57.4%	56.3%	49.2%
MOBILITY	3	0.2%	0.4%	0.6%	0.5%	0.7%	1.2%	1.0%	0.6%
	1	90.0%	81.9%	79.6%	79.5%	79.0%	74.3%	75.0%	81.8%
	2	9.3%	16.8%	19.1%	18.3%	19.6%	23.7%	23.1%	16.9%
SELF-CARE	3	0.7%	1.3%	1.2%	2.2%	1.4%	2.0%	1.9%	1.4%
	1	71.4%	56.0%	54.6%	51.5%	49.6%	47.2%	48.8%	57.2%
	2	25.0%	38.2%	39.2%	40.8%	43.3%	44.8%	44.4%	36.9%
USUAL ACTIVITIES	3	3.6%	5.8%	6.2%	7.7%	7.2%	8.1%	6.8%	5.9%
	1	50.9%	36.3%	33.2%	31.3%	30.2%	30.0%	30.5%	37.4%
	2	42.7%	53.0%	55.8%	58.0%	58.6%	58.3%	57.8%	52.8%
PAIN / DISCOMFORT	3	6.4%	10.7%	11.1%	10.7%	11.2%	11.7%	11.6%	9.8%
	1	68.2%	66.2%	64.4%	64.6%	64.8%	63.0%	64.3%	65.6%
	2	28.6%	30.9%	33.0%	32.3%	32.4%	33.2%	32.6%	31.3%
ANXIETY/ DEPRESSION	3	3.2%	2.9%	2.6%	3.2%	2.8%	3.7%	3.2%	3.1%
Patients with complete questionnaires*		3,243	1,578	1,616	1,490	1,384	1,041	1,262	3,380
Average time (years) spent in the trial (SD)		10.60 (2.84)	16.92 (2.74)	17.89 (2.78)	18.85 (2.76)	19.87 (2.75)	20.82 (2.83)	21.13 (2.75)	16.70 (4.88)

* 8 questionnaires were returned partially completed and these were not used in the analysis.

It is important to note that mean quality of life was seen to decline over time, whether or not patients experienced a major health-related event, as shown in Table 2.5.

Table 2.5: Average EQ-5D tariff for patients with and without complications by questionnaire round

Averages per questionnaire round	1	2	3	4	5	6	7	TOTAL
Tariff (SD)	**0.76** (0.27)	**0.69** (0.31)	**0.67** (0.31)	**0.66** (0.31)	**0.66** (0.30)	**0.64** (0.32)	**0.65** (0.31)	**0.69** (0.30)
Percentage of questionnaires with Tariff=1	36.1%	27.6%	23.6%	22.1%	21.3%	20.5%	21.5%	26.6%
Average Tariff (SD) in patients with **no complications**	0.79 (0.25)	0.73 (0.29)	0.71 (0.29)	0.69 (0.30)	0.68 (0.30)	0.67 (0.31)	0.68 (0.31)	0.72 (0.29)
Average Tariff (SD) in patients with at **least one complications**	0.71 (0.30)	0.62 (0.33)	0.60 (0.32)	0.61 (0.32)	0.61 (0.30)	0.57 (0.32)	0.59 (0.32)	0.64 (0.32)

Individuals who never incurred complications reported a natural deterioration of their QoL which may be attributable to ageing. The average utility for diabetic patients with no history of complications is 0.72 and that with complications is 0.64. The difference of 0.08 between these levels is often regarded as the impact on QoL when a patient experiences a complication. Ideally however, the appropriate measures of *change* should be based on the average difference in measures of HRQoL at two or more points in time t collected on the same patients before and after the health state of interest occurs (i.e., exploiting the longitudinal structure of the data).

In order to understand the degree of heterogeneity present in our sample, utility values are decomposed into two parts: variation across people, defined as *between* variation, and variation across the questionnaires of a given person, defined as *within* variation (Table 2.6). Between and within variance do not sum up because we have an unbalanced panel. The reported standard errors indicate that the variation in utility across patients is greater than the observed variation within a patient over time.

If we were to draw two patients randomly from the data, the difference in utility is expected to be almost twice the difference in utility for the same patient in two randomly selected years. If the heterogeneity across patients is correlated with the likelihood of having an event, then the results of cross-sectional analysis such as pooled OLS will be biased. This will be discussed in more detail in Section 2.4.

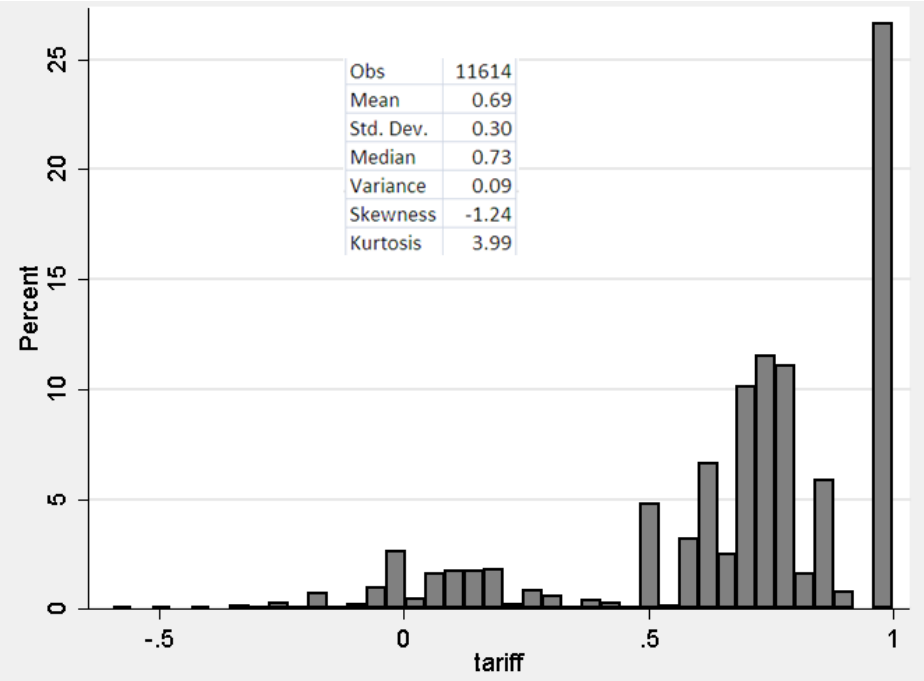
Table 2.6: Between and Within Decomposition of Variation

	Mean tariff	Standard deviation
overall	0.69	0.3
between		0.27
within		0.16

The average age of a patient answering the survey is 67 (age of the youngest patients being 32 and the oldest 92). In order to have a benchmark for comparison, the UK population tariff in a age bracket similar to that of the sample average (60-69 years of age) is 0.81 for women and 0.78 for men (Kind, Hardman et al. 1999).

Figure 2.4 shows two important characteristics of the tariff: probability mass accumulating at the upper boundary (approximately 27% of observations in this population report full health) and lack of continuity in the -0,594 to 1 range, which naturally occurs due to the amalgamation of discrete data. That there are a high proportion of people who rate themselves as having perfect health is not unlikely (Health Survey for England, 1998). A number of measures of health-related quality of life are subject to ceiling effects.

Figure 2.4: The distribution of EQ-5D tariffs

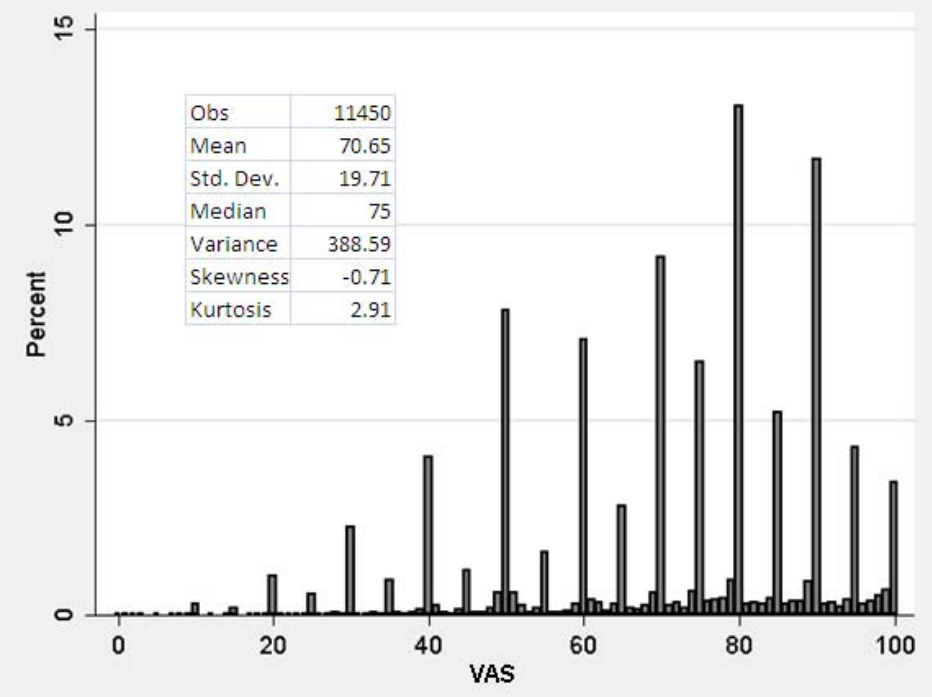


The VAS

Patients were also required to state their overall health level using a visual analog scale (VAS) by indicating a position along a continuous vertical line running from 0 to 100. Not all patients completed this part of the questionnaire; of patients who fully completed the EQ-5D part of the questionnaire 164 did not proceed to the VAS.

The VAS range of 0 to 100 indicates the worse imaginable state and the best imaginable state of health, respectively. Numbers are shown on the scale for multiples of 10. Large ticks mark multiples of 5 and small ticks every point scale (see page 2 of the valuation form in Appendix A2. 2). As shown in Figure 2.5, this nudges the responder into marking multiples of 10 more often than other numbers along the scale.

Figure 2.5: Distribution of self-reported health



The number of respondents, average VAS values, and the percentage of patients self-reporting perfect health per questionnaire round are shown in Table 2.7.

Table 2.7: Average VAS for patients with and without complications by questionnaire round

Averages per questionnaire round	**Q1**	**Q2**	**Q3**	**Q4**	**Q5**	**Q6**	**Q7**	**Total**
Patients with complete answers	3,192	1,551	1,601	1,469	1,358	1,030	1,249	3,357
VAS (SD)	**73.55** (19.05)	**71.19** (18.93)	**70.08** (19.57)	**69.60** (19.96)	**68.80** (20.00)	**68.05** (20.27)	**68.67** (20.54)	**70.65** (19.71)
Median	78	75	75	75	72	70	74	75
Percentage VAS=1	6.5%	3.0%	3.1%	1.8%	1.5%	1.7%	1.9%	3.4%
Average VAS (SD) in patients with **no complications**	75.36 (18.46)	73.95 (18.15)	72.90 (18.97)	72.42 (18.96)	71.61 (19.14)	70.16 (19.97)	75.36 (18.46)	**73.95** (18.15)
Average VAS (SD) in patients **with at least one complications**	70.26 (19.66)	66.22 (19.29)	64.80 (19.60)	64.27 (20.70)	63.55 (20.52)	63.84 (20.23)	63.60 (20.82)	**66.13** (20.18)

There is a high correlation between tariffs and VAS scores. For ease of comparison, the scale of the VAS is transformed into a 0-1 range. The OLS regression

coefficient reported in Table 2.8 shows the one-to-one correspondence and an expected slightly negative constant given the different lower support across these two measures. The VAS is however not a preference-based measure and so it is important to recognize that two individuals with identical self-assessed health status as measured by the EQ-5D instrument may value their health quite differently.

Table 2.8: OLS regression – VAS on tariff

Dependent: Tariff	Coeff.	SE
VAS	0.986**	(0.011)
Constant	-0.002	(0.008)
Observations	11450	
R-squared	0.421	

** p<0.01, * p<0.05

Complications

There were a total of ten non-fatal clinical categories in the UKPDS. In this analysis, I examine how quality of life is affected by six of these: MI, IHD, stroke, heart failure, amputation, and blindness. The remaining four events are excluded either because there were insufficient numbers of events for reliable analysis (i.e., renal failure= 38 cases), or because they were defined by the occurrence of treatment for a clinical condition (i.e.., retinal photocoagulation, cataract extraction), making it impossible to measure the disutility of the problem itself. Of course an amputation is a treatment too as it causes a loss and as such has a negative impact on QoL; however it is also a good proxy for the underlying cause (foot ulcers and gangrene for which we do not have data concomitant to the questionnaire administration), as there is virtually no patient with the underlying problem that does not undergo an amputation. The

problem with retinal photocoagulation cataract extraction and vitreous haemorrhage however is different because we also care about blindness as an endpoint. An untreated cataract or retinopathy can lead to blindness. Vitreous haemorrhage causes temporary blindness. The problem of identification arises because we are not able to distinguish between those who received the treatments and did not go blind, those who became blind despite getting the treatment, and those who were untreated despite having the underlying condition and did not go blind. Data on eye problems with an indication of the eye in question (right or left) were not collected for all patients throughout the duration of the study. For all other events the associated coefficient measures the impact on utility whether or not the person has obtained treatment as it averages those with a treatment and those without.

Approximately 32.7% of patients who answered at least one questionnaire had no complications (1,105 out of 3,380 participants); 37.8 % (1,278) had one or more micro-vascular complications (defined in the trial as blindness and other eye related disease, amputation and kidney failure); 29.9% (1,011) had macro-vascular complications (MI, IHD, Stroke and heart failure); and 13.1% (443) had both types of complications.

Table 2.9 shows the number of participants with one or more complications. For instance, there are 323 patients experiencing at least one MI, 47 of which also have IHD, 25 of whom have stroke, and so on. Each column shows how many of them also have some other complication. Because the correlation matrix is always symmetrical only the lower triangle is reported in Table 2.10, corresponding to all the possible pairings that can be derived from all the clinical endpoints. The diagonal of a

correlation matrix (i.e., the numbers that go from the upper left corner to the lower right) always consists of ones, as these are the correlations between each variable and itself. There is a weak correlation across diabetes complications possibly due to lack of power. Because according to UKPDS clinicians, eye problems culminating in temporary or permanent blindness are markers of disease progression in T2DM, there was a prior expectation of a higher degree of association between blindness in one eye (defined as at least 3 months with LogMAR>1) with the other macro-vascular complication, which did not materialize.

Table 2.9: Number of participants with one or more complications

Events	**MI**					
Myocardial Infarction	323	**IHD**				
Ischaemic heart Disease	47	356	**Stroke**			
Stroke	25	24	159	**HF**		
Heart Failure	11	9	12	134	**Amp**	
Amputation	9	12	8	7	65	**Blind**
Blindness in 1 eye	21	27	14	14	7	184

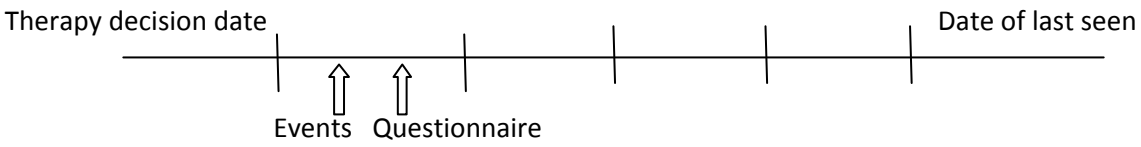
Table 2.10: Correlation matrix

Events	**MI**					
Myocardial infarction	1	**IHD**				
Ischaemic heart disease	0.05*	1	**Stroke**			
Stroke	0.04	0.04	1	**HF**		
Heart failure	-0.02	-0.03	0.04	1	**Amp**	
Amputation	0.02	0.02	0.08*	0.04	1	**Blind**
Blindness in 1 eye	0.02	0.03	0.03	0.05*	0.03	1

*P-value<0.05

Constructing the data set

If for each patient we could place the questionnaires within equally spaced temporal bins denoting the time elapsed since the subject entered the trial, the data would look as follows:



The problem with this approach is that events need to precede questionnaires in order for one to explore the question of whether or not clinical events influence questionnaire responses. Complications, however, occur at any point in time; before or after a patient answers a questionnaire. Thus a dataset needed to be constructed, which would help with the directional identification of causality. In order to construct time intervals that would allow for an accurate causal story, I define a *panel year* as comprising the 365.25 days immediately preceding any given questionnaire. In this way, every patient has a unique panel year. This way of organizing the data, however, does not address the six year gap between the first and second questionnaire. Whether or not we choose to add the 1996/7 information would depend on the model we want to estimate. If we only use contemporaneous variables (no lags), then we can use all the data available. This could be restrictive; especially if we hypothesize that the history of each individual's utilities plays a role in the contemporaneous wellbeing of a patient. To include a lagged dependent variable, we need to instrument it with an exogenous variable in order for the estimator to be unbiased. A reasonable instrument is the use of further lags or differences in lags, which imply at least three questionnaires per patients. Those with lesser observations would simple not be included in the sample, which would imply excluding approximately 45% of the sample. This is an example of how commonly a data set both constrains and informs the questions of interest.

2.4 The Model

Different ways of looking at a patient's history

QoL questionnaires are concentrated at the end of a patient's time in the trial and they are preceded by as much as 24 years of event history. This plethora of information allows us to ask a number of interesting questions, for example, does history matter? And if so, in which way is history an important predictor of current utility?

The questionnaires ask the patients about their current QoL. In cases where the probability of having an event at a particular time point is rather small, to formally account for a patient's history becomes particularly important. From a specification perspective, we are looking to approximately 10% of people having a complication. This may result in not enough power if we divide time/history preceding the questionnaire into overly small intervals. At the same time, in the case of acute events, there may be intra-period changes in quality of life that occur within surveys. So for example, the acute impact of a myocardial infarction on QoL, or the impact of temporary blindness in one eye, is unlikely to be captured for many patients if there is a long time lag between surveys.

In the empirical analysis, the time intervals of questionnaire administration set by the trial are used. These occurred yearly in accordance with the needs of cost effectiveness models. It is, however, important to look at the evidence on rehabilitation periods in order to know, for example, how long it takes for a patient to get accustomed, if at all, to a life changing state event such as amputation before returning to his or her original utility levels.

Figure 2.6: Average Utility four trimesters before and after a complication

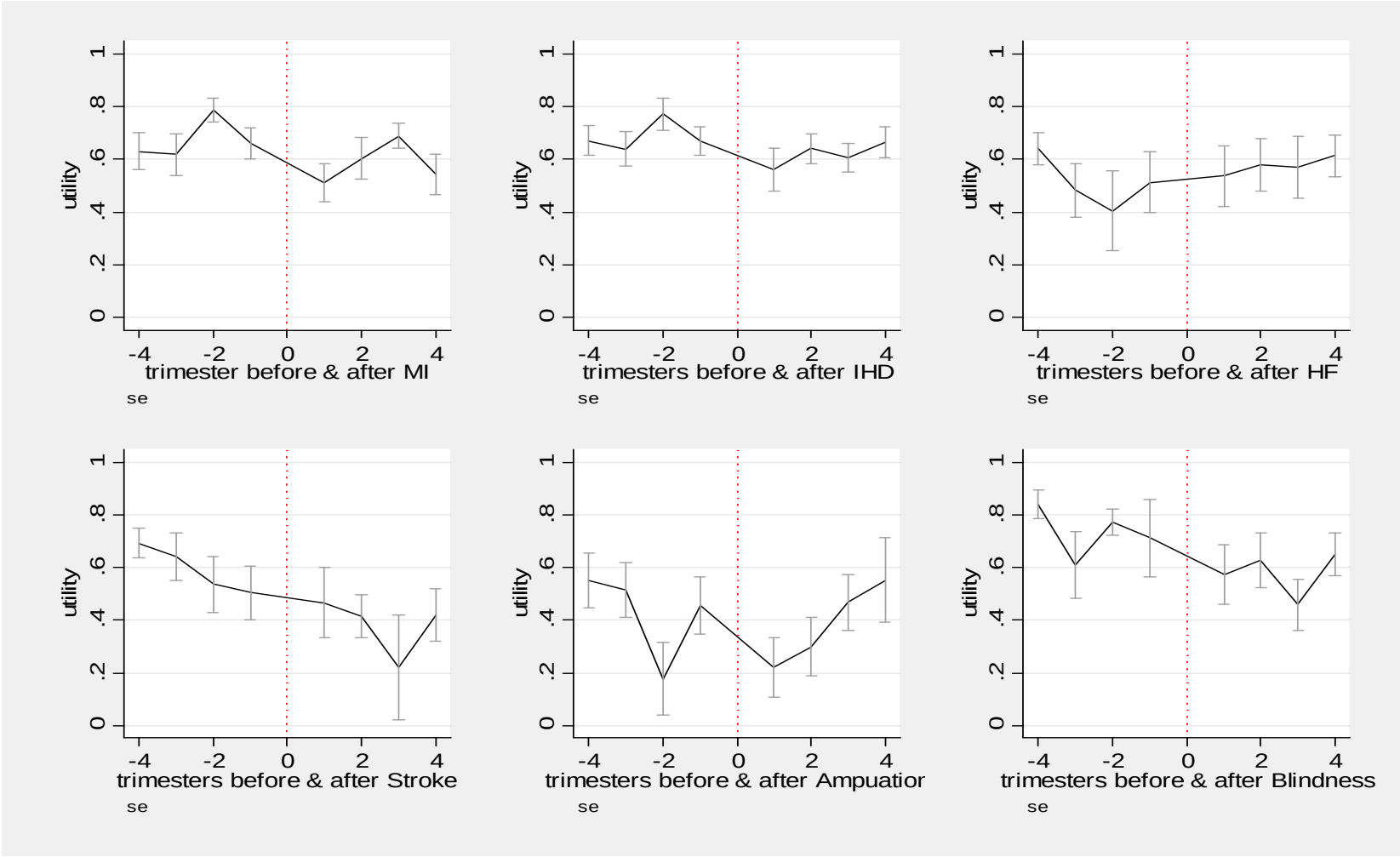


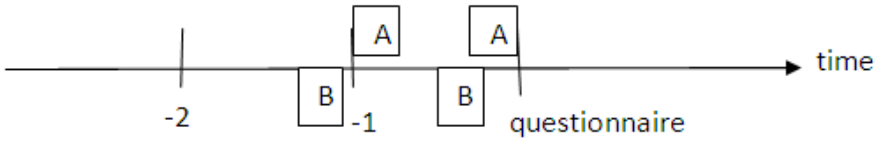
Figure 2.6 shows, albeit with a great degree of noise, the average utility for 664 patients who had at least one event either within or after a year of completing a QoL questionnaire. In it the vertical red line centres the time of the event at point zero. The patient reporting a tariff before and after an event may often not be the same. It is interesting to notice to the right of the vertical line that there is some evidence of a gradual increase in QoL post complications. Average tariffs after most complications revert to the pre-complication levels after a year. One should also note the gradual deterioration in QoL during the months preceding an event.

Clarke et al. (2002) tried to differentiate between the short and the long run impact of complications on utility by reporting the average utility of patients suffering a clinical event within the 365 days preceding the completion of a questionnaire and also in any year before the year preceding the questionnaire. This way of representing history has some limitations, not only because the definition of time is arbitrary but also because duration no longer enters into play for events occurring more than a year before a questionnaire. Events in the past, irrespectively of how recent, up to one and a half to ten years are weighted equally. As with these authors I find no significant difference between these two time periods; the only complication for which I find significant differences, using a likelihood ratio test is MI. In all other case, where differences were not found, the two dummy variables were combined, leading to the assumption that the effect of complications on utility does not vary over time.

The dichotomous representation of the history of events in a given interval might also not represent what happens to patients. For instance, similar patients may

appear to have a different history of events purely because their events fall into different time bins.

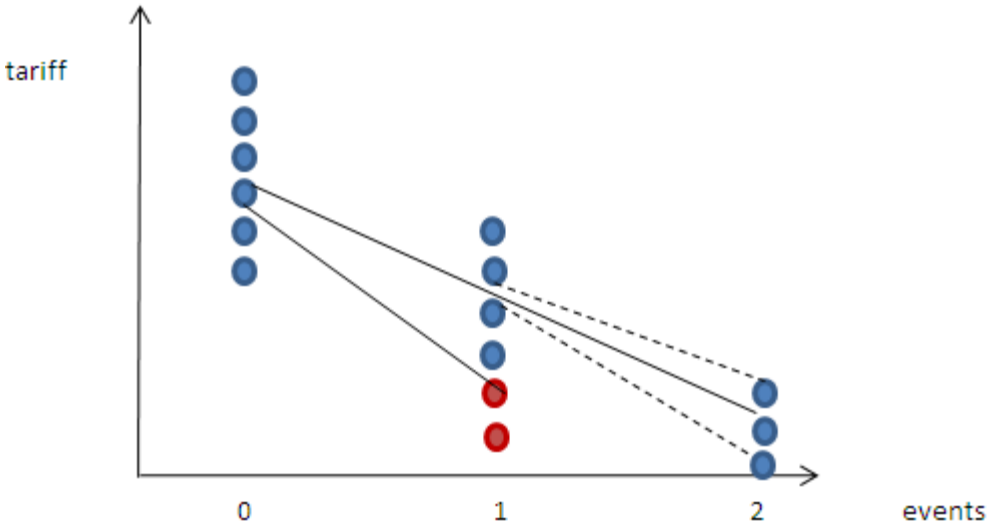
Consider this example:



Patient A and B are very similar insofar as they both have two events at equal intervals. If we use dummies, patient B appears to have more events that patient A (one in the year before a questionnaire and other previous to that), while both have had the same number of episodes in the course of their lives. Thus the dichotomous specification is subjective to the panel dimension used. Counts do not sacrifice any information on the number of complications in the data. However, the problem with estimating counts is interpretability. The marginal effect in counts represents the average number of events in the sample. Suppose events have diminishing marginal effects on QoL. The tariff scale is bounded from below. This means that after a certain point individuals cannot be made any worse off – a count explanatory variable will give a weighted average of the first, second and subsequent events, as the number of events increases, the marginal effect of each additional event is smaller in absolute terms. Another issue I considered was that the presence of additional events may not necessarily be linear. Scenarios can be conceived under which these have a steeper marginal effect (more events of the same kind may imply a more serious condition). It is likely that multiple events will affect utility more than if we only look at the question "has the patient *ever* suffered an event?", but they could flatten the marginal effect on quality of life (a person having recurrent events may learn to adapt) as shown by the dashed lines in

Figure 2.7. Using dummies to represent patients with more than one event could be a possibility, but for most complications, less than 1% of patients in the sample experienced multiple events of the same kind.

Figure 2.7: Hypothetical relation between tariff and number of events



Note: the dots in the figure represent hypothetical values of a QoL tariff in a given sample. The assumption here is that the tariff is inversely correlated with the number of events. The red dots represent the coding adopted when we ask about the relation between tariffs and having at least one event instead of counts (they represent the tariff value for people with two events in this example). The lines represent the marginal impact of having at least on event and having multiple events.

Exploiting longitudinal information

Longitudinal data allows us to control for variables we cannot observe or measure such as attitudes (the ability of a patient to cope with pain) or variables that may change over time but not across individuals (their changing environment, the availability of care in a given area). Repeated observations for the same patient over time allow us to account for individual heterogeneity.

Consistency of the pooled OLS estimator requires that what is left out of the model is uncorrelated with the variables for which we have a measure. Expressed

mathematically[3], i.e. $E(e_{it}|x_{it} = 0)$ in $y_{it} = u + \beta x_{it} + e_{it}$. If this assumption fails so that x_{it} is endogenous, IV estimation can yield consistent estimates. It can be difficult however to find an instrument z_{it} for x_{it} that satisfies the same condition: $E(e_{it}|z_{it} = 0)$.

Panel data provide an alternative way to obtain consistent estimates. The following transformation is used to attenuate the correlation between patient specific unobservable time-invariant characteristics and explanatory variables (Wooldridge 2002):

$$y_{it} - \lambda\bar{y}_i = \beta_0(1-\lambda) + \beta_1(x_{it} - \lambda\bar{x}_i) + (e_{it} - \lambda\bar{e}_i) + (u_i - \lambda\bar{u}_i)$$

where $\lambda = 1 - \left[\frac{\sigma_e^2}{\sigma_e^2 + T\sigma_u^2}\right]^{1/2}$ is between zero and one. The over bar denotes the time averages. The parameter λ is estimated through feasible GLS. Pooled OLS is obtained when $\lambda = 0$, and fixed effects (FE) is obtained when $\lambda = 1$. The random effects (RE) specification subtracts a fraction of the time average, where the fraction depends on σ_u^2 and σ_e^2 and the number of time periods. When λ is close to zero, the RE estimates will be close to the OLS estimates. This would be the case when the unobserved time invariant effect u_i is relatively unimportant. If, however, σ_u^2 is larger relative to the variance of the time varying error σ_e^2, $\hat{\lambda}$ will be closer to unity. The underlying assumption is that something within the individual may impact or bias the outcome variable of interest and one needs to control (remove) this. When using fixed effects, however, it is not possible to estimate coefficients for variables that do not vary within individuals such as gender, race, etc. One may choose to overcome this problem by

[3] Where e_{it}=error term, "what is left out of the model" *vis a vis* i (people) and t (time) ; x_{it} = the covariates of interest; y_{it}= QoL (dependent variable)

performing different analyses for subgroups (males and females), which would be equivalent to running a model with interaction effects. A caveat here is that while one may be addressing bias with a fixed effects specification, respondents with only one observation do not contribute to the calculation of the coefficients.

In order to know whether FE or RE is more appropriate in the presence of longitudinal information, researchers often use the Hausman test. The null hypothesis is that the individual effects are random, i.e., the time invariant error term is uncorrelated with the regressors. Under the null the RE and FE estimators should be similar as both are consistent. Under the alternative, these estimators diverge. The problem with this test, however, is that it does not allow the user to compare FE and RE models that have been adjusted for intergroup correlation through clustering, i.e., models with robust variance estimates. The Hausman test assumes that the error term is homoskedastic which is often not the case with a limited dependent variable. To overcome this difficulty one can use Wooldridge's (2002) test, an alternative to the Hausman test, which is robust to heteroskedasticity and at the same time provides consistent coefficients if the strict exogeneity assumption does not hold (i.e, where $u_i \perp x_{it}$ is rejected). This is carried out through a regression on de-meaned time-varying regressors:

$$y_{it} - \hat{\lambda}\overline{y_i} = (x_{it} - \hat{\lambda}_i\overline{x_i})\beta' + (x_{it} - \overline{x_i})\delta + \varepsilon_{it}.$$

The null hypothesis is then H_0: $\delta = 0$ in the auxiliary OLS regression. When a panel is unbalanced, that is, when the number of time periods is not the same for all individuals, λ_i is calculated for each patient. The test strongly rejects the null hypothesis [F(9, 3379) = 119.080; Prob > F = 0.000]. In other words, there are

unmeasured patient-specific characteristics that are correlated with both the outcome of interest (tariff) and the likelihood of complications that would render an estimator (RE or OLS) that does not use within differences (such as FE) biased.

The issue of Non-Response

We have missing questionnaires for some patients in the sample. Although the estimation mechanics remain unchanged whether or not we have observations across all seven questionnaires for all the patients in the sample, we worry whether the unbalanced panel has underlying selection effects on unobservables. Provided the reason for missing data is not correlated with the idiosyncratic error term, the unbalanced panel causes no problems. If however the reason a patient ceases to complete a questionnaire is correlated with the idiosyncratic error, those unobserved factors that change over time and affect utility, then the resulting sample selection can indeed cause biased estimators.

Differences between respondents who answered one EQ-5D survey during the period of follow-up and those answering multiple questionnaires were investigated. In order to make these two groups comparable, all summary statistics correspond to the time of the first questionnaire administered between 1996 and 1997. Table A2. 6 in the Appendix indicates that there are some differences between these two groups. Those who answer multiple surveys tended to be younger (61 vs. 63 years), had slightly shorter durations of diabetes, had higher utility values and, on average, a lower number of macro-vascular complications than singletons at the time of the first questionnaire. Table A2.7 shows the same pattern for tariffs at subsequent

questionnaire rounds. If patients differ in terms of observable characteristics we might[4] face a problem of external validity if we do not control for these differences.

Attrition is an all absorbing state, indicating permanent loss of the patient from the sample. Non-response is a temporary phenomenon as the individual rejoins the survey at subsequent rounds. As Wooldridge (2002) points out, the sequential nature of attrition makes *first differencing* a natural choice to remove unobserved effects, when these are present. Consequentially, instead of looking at attrition, which in the majority of cases corresponds to death[5], I attempt to model non-response hereafter.

I use *(sample) selection* as a general term to refer to any process that governs whether an individual submits a complete EQ-5D questionnaire, and thus has a corresponding tariff. I make the general assumption that the selection processes is *non-random*, in the sense that it may be correlated with the complications under study. I thus use a selection correction model for non-response while treating attrition as random.

For simplicity of exposition, let us assume that the relationship between complications, x, and EQ-5D tariff rates, y, is as follows: $y = \beta x + \epsilon$, where ϵ encompasses all unobserved determinants of tariff rates. If we further assume that *in the population as a whole* there is no correlation between the onset of events, x and the unobservables ϵ, β is interpretable as the causal parameter of interest unless one of these two conditions presents itself (Heckman 1979):

[4] It is worth noting that different groups could still have homogeneous "treatment effects". Consider for example the relationship between being hit by a train and dying. The coefficient is one regardless of who is hit by the train but the incidence across characteristics may differ. A drunkard may be more likely to be hit by a train than a sober person.

[5] Death is a perfect prediction of selection and as such its corresponding coefficient should approach infinity.

Condition 1. The explanatory variable(s) of interest, x are correlated with the probability of being observed in the sample.

Condition 2. The *unobserved* determinants of the sample selection process are correlated with the unobserved determinants of the outcome of interest, ϵ.

To illustrate this more clearly, consider the following latent-variable model for sample selection:

$$s^* = \boldsymbol{x}'\boldsymbol{\alpha} + \mathrm{w}'\gamma + u\ ; \qquad u \sim N(0,1)$$

s^* is not observed but indicates the patient's willingness to answer a questionnaire. The researcher is only able to observe whether or not the patient has answered a questionnaire

$$s = \begin{cases} 1 \text{ if } s^* > 0 \\ 0 \text{ if } s^* \leq 0 \end{cases}$$

$\boldsymbol{x}'\boldsymbol{\alpha}$ denotes a matrix of covariates and their associated coefficients. This matrix contains the same covariates of interest which will be analyzed in the equation of interest. w is a vector of variables that are assumed to be excluded from the outcomes equation. The importance of an exclusion restriction is twofold: the restriction (hypothesis) becomes testable, and may make a simultaneous equation system identifiable as it no longer relies purely on functional form. In cases where the same variables are used in both the selection and outcomes equation, the inverse Mills ratio may be collinear to other variables, especially if they are dichotomous ones.

For FE the reader should assume subscripts *i,t* apply corresponding to *1=1...N* individuals and *t=1,...T* questionnaires. OLS instead assumes each observation is an independent draw so the relevant subscript is the patient questionnaires j=1...NT.

Condition 1 implies that $\alpha \neq 0$, while condition 2 implies that $Corr(\epsilon\ u) \neq 0$.

By way of example, suppose that $\alpha < 0$ and $Corr(\epsilon\, u) > 0$. This is perhaps the most intuitive case, meaning that complications reduce the likelihood of submitting a questionnaire, and individuals who experience higher tariffs are more likely to participate. An individual with complications x will only submit a questionnaire if they experience unobservables u such that $s^* > 0$. This implies that $u > -x'\alpha$.

For a scalar $x \in (0,1)$ and $\alpha < 0$ (i.e., a complication makes a patient less likely to submit a questionnaire), the set of possible values of u, *conditional on a questionnaire being submitted*, is illustrated together with their distribution for the two values of x in Figure 2.8 (here w=0 for simplicity).

Figure 2.8: Examples of distribution of the error in the selection equation without and with a complication

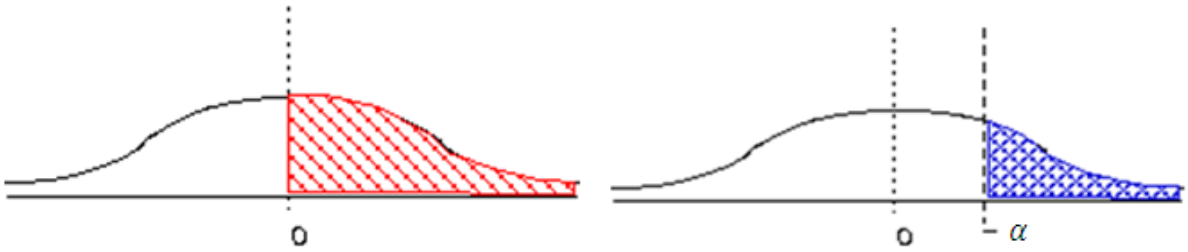


Fig. 2.8a. Distribution of u for individuals with $x = 0$

Fig 2.8b. Distribution of u for individuals with $x = 1$

As illustrated, by this example, individuals with complications must experience a higher unobserved 'shock', u, to the likelihood of participation than individuals without complications, in order to be observed in the sample. Consequently, the expected value $s = 1$ is a function of x, even though x and u may be uncorrelated in the population as a whole. Heckman's insight into a parametric solution to the selection problem stems from the observation that when the u are normally distributed, such that the selection model has the probit form, then this conditional expectation is given by the inverse of the Mills ratio.

Why does the fact that the expectation of u depends on x in the sampled population create a source of bias? This depends on *Condition 2*, namely, that u is correlated with ϵ. Under this assumption, which is testable in the Heckman framework, the expectation is $E[\epsilon|u] \neq E[\epsilon]$. Therefore, if the expected value of u in the selected population is a function of events x, the expected value of ϵ will be a function of x in the population we observe as well. As Heckman observed, this can be thought of as a classic case of omitted variable bias. Absent any corrective steps, the correlation between x and ϵ in the sample will bias estimates of the coefficient of interest, β. If α is negative and ϵ, u are positively correlated, then this bias will be upward (i.e., towards zero if the true β is negative, or away from zero if the true β is positive).

Benefitting from the panel dimension, we are able to interpret the observed differences between those who never incur an event and those that do incur them by analysing a patient's participation in the survey and test for sample selection. The question here is: are the results we obtain from a selected sample of responders intrinsically different from the trial population? The use of longitudinal data also resolves (or reduces) the magnitude of omitted (unobserved) variables correlated with the explanatory variables of interest. Non-response in fact could be resolved by eliminating the time invariant error term. By utilizing the information on both inter-temporal and individual entities, one is better able to control for both the effects of missing and unobservable variables.

The canonical Heckman model applied to longitudinal data

Wooldridge (1995) proposed a fixed effects (FE) model in which the unobserved component is assumed to be correlated with the observed explanatory variables. The advantage of the Wooldridge procedure is that it is relatively computationally simple. His method requires a standard Probit for the selection equation, but in order to estimate a probit in a longitudinal setting the regression is conducted for each time period. This overcomes the problem that fixed effects cannot be conditioned out of the likelihood in a probit model. Thus, if a patient completes all seven questionnaires, for that individual there will be seven regressors for each of the explanatory variables. We allow for the possibility that:

I -- $Corr(u_{it}\ \epsilon_{it}) \neq 0$, which is testable

II -- $E(x_{it}\ \epsilon_{it}) = 0,\ E(u_{it}\ \epsilon_{it}|s_{it}) \neq 0 \rightarrow \gamma IVMR$

The selection mechanism is always observed as we know who was sent a questionnaire and when, and we know when and if it was completed. The selection equation in the panel context is then:

$$s_{it}{}^{*} = \text{constant} + \boldsymbol{x_{it}}'\boldsymbol{\alpha} + x_{i1}\alpha_1 + \cdots x_{i7}\alpha_7 + \ \mathrm{w}_{it}'\gamma + \xi_i + a_{it}$$

The error term is composed of two terms $u_{it} = \xi_i + a_{it}$, ξ_i is a time-constant unobserved effect:

$$s = \begin{cases} 1 \text{ if } s^{*} > 0 \\ 0 \text{ if } s^{*} \leq 0 \end{cases}$$

$$\hat{u}_{it} = s_{it} - \widehat{\boldsymbol{x}'\boldsymbol{\alpha}}$$

$$\ddot{y}_{it} = \ddot{x}_{it}\beta + \rho\ddot{u}_{it} + \text{error}$$

Where $\ddot{u}_{it} = \hat{u}_{it} - \frac{1}{T}\sum_{r=1}^{T} s_{ir}\hat{u}_{ir}$

A serial correlation and heteroskedasticity robust standard error is computed for each ρ.

If H_0: $\rho = 0$, then non-response is not important.

If, on the contrary, we reject the null hypothesis, then we proceed to carry out the full specification as in Procedure 4.2 in Wooldridge (1995).

For each t=1,2...7, we estimate $\Pr(s_{it} = 1|x_{it}) = \Phi(\boldsymbol{X}'\,\boldsymbol{\alpha})$ using standard probit. For $s_{it} = 1$ we compute $\hat{\lambda}_{it} = \lambda(\boldsymbol{X}'\,\boldsymbol{\alpha})$ where $\boldsymbol{\lambda}(.)$ denotes the inverse Mill's ratio.

$$y_{it} = \text{constant} + \boldsymbol{x_{it}}'\boldsymbol{\beta} + x_{i1}\beta_1 + \cdots x_{i7}\beta_7 + INMR_{i1}\delta_1 + \cdots + INMR_{i7}\delta_7 + \epsilon_{it}$$

It is important to note here that *s* and y are part of the data generating process. The inverse Mill's ratio is not. Because the IVMR are calculated variables we need SEs that reflect this uncertainty. Instead of constructing analytical standard errors, I use bootstrapped standard errors. The standard deviation of the bootstraps is the estimate of the SE for the coefficient. The above equation is read simply as $y_{it} = \boldsymbol{x_{it}}'\boldsymbol{\beta}$ and all other variables are treated as controls and are not of intrinsic interest. The interest only lies insofar as the other variables affect the estimates of the betas.

Although the validity of an exclusion restriction cannot be tested on its own but only in relation to exclusion restriction or by comparing it with results obtained without an exclusion restriction, one can think of it as an instrumental variable (IV) of the Inverse Mills ratio. The variable has to influence selection but it must not have any direct impact on the outcome of interest, which is utility.

The interaction between location and time is used in this analysis as the exclusion restriction[6]. One may argue that location alone and utility could be linked (i.e., that people in Scotland could be perceived to be more unhealthy than the people of England). If that is the case and we were looking into a level's specification, then location would be a bad exclusion restriction, and the results would be questionable even if location and utility were indeed related. However, by exploiting the variation through time within location, the FE specification picks up any systematic differences within location.

It is possible that selection could be driven purely by patient specific characteristics. If that is the case, a simple fixed effects model, which gets rid of the time-invariant noise, could be used. We test this hypothesis in section 2.5.

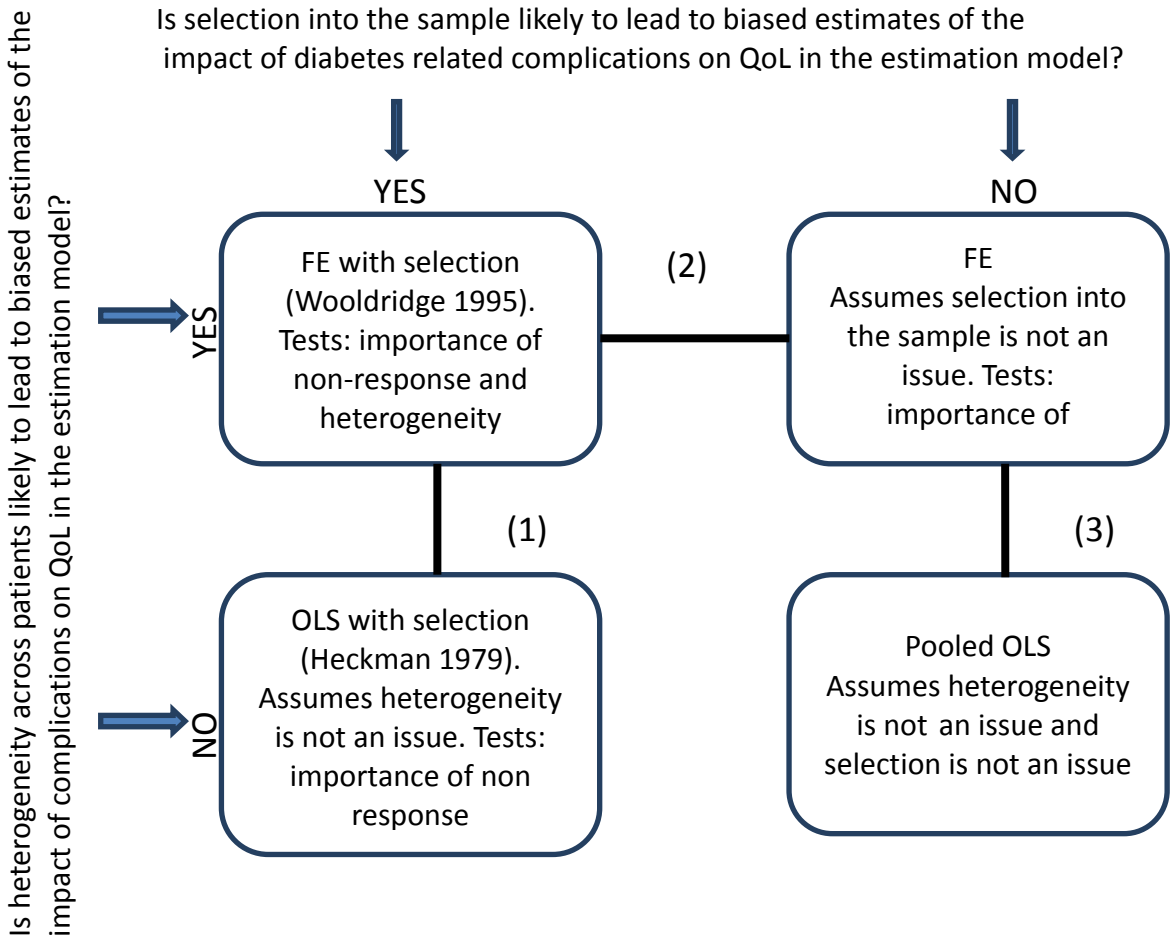
Estimation Strategy

A diagram may best explain the modelling strategy adopted. We start in the south-east quadrant of Figure 2.9, by testing FE with attrition (Wooldridge 1995) because it demands the smallest possible amount of restrictions. This model is the most robust under the weakest assumptions as it takes into account both the (1) selection process as well as (2) possible bias due to patient specific unobserved characteristics. Section 2.5 shows that non-response matters marginally (in the Heckman model) but that differences in participation rates are accounted for by FE,

[6] Joint test that all locations are equal to zero: F(22, 28821) = 4.54. Prob > F =0.0000 in the equation explaining selection through the use of the proposed exclusion restriction.

which drives us to the west of Figure 2.9 and ultimately to the choice of a simple FE model.

Figure 2.9: Schematic Approach to the Estimation Strategy



Notes: the estimation strategy tests for (1), (2) and (3) is as follows:

(1) Wald test of indep. eqns. (rho = 0): χ2(1) = 1.18; Prob > χ2 = 0.2775 in the Heckman Model. The inverse Mills ratio of the FE Wooldridge model is not statistically significant (P-value=0.669).

(2) If $H_0: \rho = 0$, then non-response is not important in $\ddot{y}_{it} = \ddot{x}_{it}\beta + \rho\ddot{u}_{it} +$ error where $\ddot{u}_{it} = \hat{u}_{it} - \frac{1}{T}\sum_{r=1}^{T} s_{ir}\hat{u}_{ir}$ (Appendix A2. 9)

(3) A Hausman test robust to heteroskedasticity and cluster effects strongly rejects the null hypothesis of equal coefficients between FE and Pooled OLS for both Model

specifications presented in Table 2.12 and Table 2.13. Full specifications are available in Appendix A2. 12.

2.5 Estimation and Results

Following the estimation strategy set out in Figure 2.9, the results of the Heckman and the Wooldridge sample selection outcomes models are reported in Table 2.11.

Table 2.11: Results of outcomes equations in Heckman and Wooldridge Selection Models

Outcomes (Part 2) †	OLS with attrition Heckman (1979)		FE with attrition Wooldridge (1995)	
	Coeff	SE	Coeff	SE
Constant	0.866**	(0.042)	0.866**	(0.029)
Age at Q1	-0.001*	(0.001)	-0.001**	(0.000)
Q2	-0.063**	(0.011)	-0.102**	(0.022)
Q3	-0.073**	(0.010)	-0.056**	(0.021)
Q4	-0.086**	(0.011)	-0.051*	(0.023)
Q5	-0.089**	(0.012)	-0.052*	(0.026)
Q6	-0.107**	(0.013)	-0.063*	(0.029)
Q7	-0.099**	(0.013)	-0.036	(0.032)
Male=1	0.073**	(0.010)		
Asian	-0.032	(0.019)		
events				
MI (year before)	-0.112**	(0.038)	-0.110**	(0.026)
MI (prior history)	-0.037*	(0.018)	-0.014	(0.017)
IHD	-0.072**	(0.017)	-0.019	(0.016)
Stroke	-0.180**	(0.032)	-0.165**	(0.023)
Heart Failure	-0.175**	(0.034)	-0.133**	(0.027)
Amputation	-0.182**	(0.042)	-0.131**	(0.027)
Blindness in 1 eye	-0.034	(0.023)	-0.020	(0.021)
location	See Appendix A 2.8 and A 2.9 for complete list of coefficients on 23 locations			
Observations	22274		Observations	22274
			Number of pts.	4252
IMR (λ)	-0.011	(0.010)	-0.004	(0.006)
rho (ρ) ‡	-0.039	(0.036)		
exclusion restriction	location*time		location*time	

†Part one (the Selection equations can be found in the Appendix, A 2.8 and A 2.9)
‡Wald test of indep. eqns. (rho = 0): $\chi2(1) = 1.18$; Prob > $\chi2 = 0.2775$
** p<0.01, * p<0.05
Computed SEs are cluster robust
Test of OLS with Selection and FE with Selection=F(49, 2662) = 3.03; Prob > F=0.000

Results of the selection parts are reported in Appendices A2.8 and A2.10. Selection variables appear to affect the likelihood of being censored out of the sample but as the estimated ρ is not significantly different from zero, we cannot reject the hypothesis of

no correlation between the error terms from the participation decision and the tariff equations (P-value=0.277). λ provides the estimated coefficient on the inverse Mills ratio. When λ is different from zero sample selection bias is present. From the estimates reported in Table 2.12 both ρ and λ indicate that the selection and outcomes equations are independent and that the OLS coefficients are no different statistically from the Heckman corrections. The inverse Mills ratio in the FE Wooldridge model is also not statistically significant.

Comparison of Fixed Effects and Pooled OLS estimators

The importance of accounting for between patient heterogeneity is addressed through a comparison of the FE regression models and the pooled linear regression models. As the number of events within each category of diabetes-related complications is not always sufficient in this dataset to illustrate statistically significant differences in individual coefficients, Table 2.12 shows results with the presence or absence of any complication captured by a single variable and Table 2.13 then shows results for the full set of complications. These comparisons show that the fixed effects model is preferred in all comparisons. The hypothesis of a single intercept across patients is rejected ($p<0.001$), indicating that there is a high degree of heterogeneity and that the pooled OLS could produce inconsistent estimates. Therefore, in this situation, the less restrictive FE models represent the preferred statistical model for estimating the impact of diabetes related complications on quality of life. When all complications are grouped together (Table 2.12), the coefficient is significantly lower (-0.054 versus -0.097; difference 0.042, 95% C.I.0.006, 0.078). When analysed separately, (Table 2.13),

the coefficients on specific diabetes complications are not different between FE and OLS at conventional levels of significance, but the FE estimates are consistently smaller (i.e., the effects are closer to zero) than their OLS counterparts. This suggests that the use of OLS could potentially result in over-estimates of the impact of complications on quality of life.

The FE model is also tested against a more fully adjusted OLS model (column 6, Tables 12 and 13) using all other time-invariant covariates available to us to control for heterogeneity. The results indicate that the differences between FE and OLS models are slightly reduced but remain highly significant ($p<0.001$). Amputation, stroke and heart failure, in that order have the highest effect on utility loss. We find no effect of blindness on QoL. This may be due to how blindness was defined in the trial, as LogMAR>1 in one eye for at least 3 months. As the measure relates to the worst eye, patients may be unaware of the problem as long as vision in the best eye is maintained, and may fail to realize an impairment has occurred, or may be aware of it but find that the impact of HRQOL is negligible.

Table 2.12: Comparison of Fixed Effects and Pooled OLS estimates (grouping complications)

Dependent variable:	FE		Pooled OLS		Difference (FE-OLS)	95% CI	Pooled OLS with additional covariates		Difference (FE-OLSadj)	95% CI
Tariff	Coeff.	SE	Coeff.	SE			Coeff.	SE		
Constant	0.811**	(0.005)	0.855**	(0.031)			0.853**	(0.023)		
Age at Q1			-0.001*	(0.001)			-0.001*	(0.001)		
Q2	-0.116**	(0.007)	-0.066**	(0.008)	-0.050	-0.071; -0.029	-0.071**	(0.008)	-0.045	-0.066; -0.024
Q3	-0.124**	(0.007)	-0.081**	(0.008)	-0.044	-0.064; -0.023	-0.081**	(0.008)	-0.043	-0.064; -0.022
Q4	-0.146**	(0.007)	-0.093**	(0.008)	-0.053	-0.074; -0.031	-0.097**	(0.008)	-0.049	-0.071; -0.027
Q5	-0.152**	(0.008)	-0.094**	(0.009)	-0.058	-0.080; -0.035	-0.099**	(0.009)	-0.054	-0.076; -0.031
Q6	-0.174**	(0.009)	-0.112**	(0.010)	-0.062	-0.088; -0.037	-0.116**	(0.010)	-0.058	-0.084; -0.032
Q7	-0.169**	(0.008)	-0.103**	(0.009)	-0.066	-0.091; -0.042	-0.109**	(0.009)	-0.060	-0.084; -0.036
Complication (=1)	-0.054**	(0.014)	-0.097**	(0.012)	0.042	0.006; 0.078	-0.106**	(0.012)	0.051	0.015; 0.087
Male(=1)							0.074**	(0.010)		
Asian							-0.028	(0.019)		
Manual worker (comparator: Professional and managerial)							-0.054**	(0.012)		
Unskilled worker (comparator: Professional and managerial)							-0.068**	(0.014)		
location							See Appendix 2.10 for complete list of 23 locations			
** p<0.01, * p<0.05					‡Test: Ho: difference in coefficients				‡Test: Ho: difference in coefficients	
Observations	11614		11614		not systematic		11328		not systematic	
Number of patients	3380		3380		Sargan-Hansen stat.= 90.44		3281†		Sargan-Hansen stat.= 80.57	
R-squared	0.121		0.046		Chi_sqr(7) P-value=0.000		0.102		Chi_sqr(7) P-value=0.000	

** p<0.01, * p<0.05

† Social class was not available for all participants.

‡Cluster robust Hausman test

Table 2.13: Comparison of Fixed Effects and Pooled OLS estimates (all complications)

Dependent variable:	FE		Pooled OLS		Difference (FE-OLS)	95% CI	Pooled OLS with additional covariates		Difference (FE-OLSadj)	95% CI
Tariff	Coeff.	SE	Coeff.	SE			Coeff.	SE		
Constant	0.807**	(0.005)	0.845**	(0.031)			0.849**	(0.039)		
Age at Q1			-0.001*	(0.001)			-0.001	(0.001)		
Q2	-0.114**	(0.007)	-0.065**	(0.008)	-0.049	-0.063; -0.035	-0.070**	(0.008)	-0.044	-0.065; -0.023
Q3	-0.123**	(0.007)	-0.080**	(0.008)	-0.043	-0.064; -0.022	-0.080**	(0.008)	-0.042	-0.063; -0.022
Q4	-0.143**	(0.007)	-0.092**	(0.008)	-0.052	-0.073; -0.031	-0.095**	(0.008)	-0.048	-0.070; -0.026
Q5	-0.149**	(0.008)	-0.093**	(0.009)	-0.056	-0.078; -0.034	-0.097**	(0.008)	-0.052	-0.074; -0.029
Q6	-0.172**	(0.009)	-0.111**	(0.010)	-0.061	-0.085 ; -0.037	-0.115**	(0.010)	-0.057	-0.083; -0.031
Q7	-0.167**	(0.008)	-0.102**	(0.009)	-0.064	-0.090; -0.039	-0.109**	(0.009)	-0.058	-0.082; -0.034
MI (year before)	-0.065*	(0.030)	-0.086*	(0.036)	0.020	-0.041; 0.082	-0.102**	(0.037)	0.037	-0.056 ; 0.130
MI (prior history)	0.008	(0.024)	-0.019	(0.018)	0.027	-0.057; 0.112	-0.038*	(0.018)	0.047	-0.012; 0.105
IHD	-0.028	(0.022)	-0.067**	(0.017)	0.039	-0.016; 0.094	-0.073**	(0.016)	0.046	-0.008; 0.099
Stroke	-0.165**	(0.035)	-0.179**	(0.030)	0.014	-0.062; 0.090	-0.193**	(0.030)	0.027	-0.063; 0.118
Heart Failure	-0.101**	(0.032)	-0.168**	(0.032)	0.067	-0.018; 0.152	-0.166**	(0.031)	0.065	-0.021; 0.152
Amputation	-0.172**	(0.045)	-0.180**	(0.040)	0.007	-0.101; 0.115	-0.201**	(0.039)	0.029	-0.088; 0.146
Blindness in 1 eye	0.033	(0.027)	-0.050*	(0.022)	0.083	-0.011; 0.177	-0.038	(0.022)	0.070	0.002; 0.139
Male(=1)							0.074**	(0.010)		
Asian							-0.033	(0.019)		
Manual worker (comparator: Professional and managerial)							-0.053**	(0.012)		
Unskilled worker (comparator: Professional and managerial)							-0.066**	(0.013)		
location							See Appendix 2.10 for complete list of 23 locations			
** p<0.01, * p<0.05					‡Test: Ho: difference in coefficients				‡Test: Ho: difference in coefficients	
Observations	11614		11614		not systematic		11328		not systematic	
Number of patients	3380		3380		Sargan-Hansen stat.= 87.51		3281†		Sargan-Hansen stat.= 76.11	
R-squared	0.131		0.082		Chi_sqr(13) P-value=0.000		0.102		Chi_sqr(13) P-value=0.000	

** p<0.01, * p<0.05

† Social class was not available for all participants.

‡Cluster robust Hausman test

Duration and interaction of quality of life impact of complications

The analysis also tested whether the effects of complications in the acute phase (within one year of an event) and longer term (more than a year after an event) differ. There was evidence that this was the case for MI but not for other complications. Consequently the quality of life impact reported for MI is separated in Table 2.13 into short-term and long-term decrements, while the impacts of all other complications can be interpreted as permanent decrements following the respective complications.

We also tested for evidence of interactions between the effects of the six diabetes related complications which might result in the combined effect of more than one complication having a greater or lesser impact on quality of life than separate effects. However, we found no evidence of significant interactions.

The FE results reported in Table 2.13 can therefore be considered as additive or independent effects on quality of life and are the preferred estimates. The results show that amputation (-0.172 SE[0.045]), stroke (-0.165 SE[0.035]) and heart failure (-0.101 SE[0.032]) have the largest impact on quality of life, followed by the short-term impact of an MI (-0.065 SE[0.030]) . We do not find any statistically significant effect of blindness in one eye (0.033 SE[0.027]) or of ischaemic heart disease (-0.028 SE[0.022]) on quality of life.

In order to have an idea of the differences in magnitudes between a cross sectional estimation and a longitudinal estimation, we compare the difference in coefficients. Not only are the coefficients statistically different but the difference is also significant from the clinical point of view. Norman et al. (2003) show that when

patients with chronic diseases are asked to identify the minimal level of real change, the estimates denoting a minimal level of real change in the QoL scale fall very close to half of a standard deviation. Using the UKPDS sample as the benchmark for the distribution, the decrements in utility will correspond to 0.015. Walters et al.(2005) found similar results and showed that the MID (minimum important difference) for the EQ-5D was between 0.011 and 0.014. VAS scores have also been used in determining the impact of complication on QoL, although these are subjective measures (results can be found in Appendix Table A2.13).

2.6 Conclusion

Observing a patient's responses over multiple time periods gives us several major advantages over a conventional cross-section. We have a larger number of data points, increasing the degrees of freedom of the estimation, hence improving the efficiency of the estimates. More importantly, we are able to answer a number of important questions that could not be addressed using cross-sectional observations alone, particularly the impact of non-response as well as the effect of a *change* in status.

The measure of change in utility is central to decisions about the use of alternative drug regimes by reimbursement agencies worldwide (e.g. NICE, PBAC, Medicare). Reliable estimates of the impact of health events on quality of life are essential when estimating the cost-effectiveness of health care interventions aimed at reducing diabetes-related complications. Reliable estimates depend not only on having

very large data sets, as the incidence rate of complications is relatively small, but also on repeated observations over time.

In this chapter, an empirical analysis was used to compare estimates of the quality of life impact of diabetes-related complications derived from cross-sectional data with estimates from longitudinal data. The hypothesis was that the cross-sectional model might be substantially flawed since patients who have had complications may be systematically different from those who do not, in both their observed and unobserved characteristics. To address this problem, I analysed a longitudinal dataset using a fixed effects model that accounted for the impact of both non-response to EQ-5D questionnaires as well as heterogeneity between patients with and without non-fatal diabetes-related complications.

Herein I also report quality of life decrements associated with six diabetes-related complications of type 2 diabetes. In so doing I draw on the clinical history of patients in the sample from the time of diabetes diagnosis onwards, as well as on quality of life data reported over multiple time points. These longitudinal data allow us to test whether or not sample selection (non-response) matters, and to control for the effects of time-invariant omitted variable bias. We find that reliable estimates of quality of life decrements depend not only on adequate sample sizes, but also on the availability of longitudinal data on complications and quality of life.

In the current analysis, and possibly more generally, failing to account for patient heterogeneity and relying on cross-sectional estimation methods is likely to overestimate the impact of complications on quality of life. The reason for this is that cross-sectional analyses attribute all differences in quality of life between those who do

or do not experience events to the events themselves, without taking into account that the two groups could differ in other aspects not adjusted for in the estimation model. In contrast, FE estimation takes into account before and after differences in quality of life in individuals who experience events, and therefore mitigates the potential for heterogeneity bias due to time-invariant patient characteristic. It is important to point out that the true magnitude of bias is unknown and any generalization of the results reported here would require a simulation study. Nevertheless, the estimates of bias in quality of life decrements are likely to be clinically important.

The current work employed linear regression methods to examine the impact of non-response and patient heterogeneity on estimates of quality of life change between health states. These methods have well developed estimators for dealing with such issues, particularly in a panel data context. It should be acknowledged that quality of life data have properties, such as being bounded above, and in the case of the EQ-5D have "ceiling effects" manifested by a significant proportion of the sample indicating no problems in any of the five domains of the instrument. Examining the impact of these data characteristics in a panel estimation framework would be an important contribution for future work, although estimators that exploit within patient variation become increasingly intractable in the context of non-linearity. In the current study the primary concern has not been with the distributional form, which mainly affects efficiency, but rather with consistent estimation of causal effects in light of the potential bias arising from non-response and patient heterogeneity.

A further interesting extension of this study would involve survey design. Given that longitudinal studies administer surveys on multiple occasions it is important to

determine when and how often to collect information on HRQoL. There are trade-offs in the design of longitudinal studies, as lengthening the period of time between questionnaires allows for more patients to experience the health state of interest, but it also increases the chance that an individual who has experienced deleterious health state will die before being surveyed. Further, in the case of acute events, there may be intra-period changes in quality of life that occur between surveys.

The results indicate that the fixed effects approach is likely to be more robust than a pooled OLS approach when estimating the "real" effects of diabetes-related complications on QoL. Indeed fixed effects models are consistently preferred when compared with the pooled OLS model. The estimated magnitude of quality of life effects in this dataset was consistently less in the fixed effects model. For example, pooled OLS indicated that the effect of ischaemic heart disease was to lower quality of life by 0.067 (0.073 with additional time-invariant controls) whereas the fixed effects model indicated a statistically non-significant reduction of 0.028. This suggests that correlations between the occurrence of diabetic complications and unmeasured or unused determinants of QoL systematically bias the coefficients in the OLS regressions downwards, and thus the impact of complications on quality of life is over-predicted. One possible explanation for this finding is that patients who at some point have a diabetes-related complication are already on a lower utility path than patients who never incur events, other things equal. The addition of further covariates in a pooled OLS model suggests that this problem might be reduced by the collection of additional variables which capture patient heterogeneity that may influence quality of life. However, quite apart from the potential costs and difficulties of obtaining such data, it

is important to note that, even with their addition to the analysis the estimates of the OLS model remain borderline statistically different from the estimates of the FE model. Moreover, the OLS model still relies heavily on strong assumptions, for example that the response of quality of life to these additional covariates is linear and that there is no heterogeneity in the impact of the additional covariates. In contrast, the fixed effects model by construction adjusts for time-invariant patient specific characteristics, and avoids the need to collect extensive patient-specific information.

How important are the potential inconsistencies in the estimates of QoL effects derived with OLS models? The literature on minimum clinically important differences in quality of life in patients with chronic disease indicates that the minimum is around half a standard deviation (Norman, Sloan et al. 2003), or approximately 0.015 in our sample. Similarly, Walters and Brazier (Walters and Brazier 2005) have suggested that the minimum important difference for the EQ-5D is between 0.011 and 0.014. In this analysis, the differences in estimates of quality of life decrements between minimally adjusted pooled OLS and FE were all in excess of such minimum important differences: 0.067 for heart failure; 0.039 for IHD; 0.020 for history of MI; 0.014 for stroke; and 0.083 for blindness in one eye.

Further empirical work using a range of longitudinal datasets in different disease areas is clearly needed to assess the generalizability of this results, but the study using this dataset suggests that repeated measures data analysed using the fixed effects approach is likely to give a more reliable estimate of the effects of health events on quality of life than cross-sectional data. FE models do, however, have the drawback of increasing the noise to signal ratio, especially when within patient variation is low.

Had data permitted, it would have been interesting to present a richer FE model alongside the adjusted OLS model, including a full set of behavioural variables. But the danger with variables that present a high degree of persistence, such as smoking, alcohol consumption, fitness levels, etc. is that these might not have been easily identifiable because of the low degree of within patient variation.

For those who only have access to cross sectional data, the question is can estimates of the marginal differences due to complications be adjusted to remove bias due to between individual characteristics? This is an important issue for further analysis, particularly as the vast majority of studies to date have involved cross-sectional rather than longitudinal data. Fruitful avenues for investigation would be the inclusion of additional control variables in OLS, the use of instrumental variables, and the use of methods such as propensity score matching.

The substantive results are broadly in line with previous estimates of the effect of diabetes related complications on quality of life, in particular the finding that amputation and stroke have the largest effect. However, the finding that blindness in one eye did not have a significant impact on quality of life does differ from previous studies (Brown, Brown et al. 1999; Sharma, Oliver-Fernandez et al. 2003; Clarke, Simon et al. 2006; Glasziou, Alexander et al. 2007; Huang, Brown et al. 2007) which have all been based on cross-sectional data. This difference may be driven by the estimation method: the difference between fixed effects and pooled OLS was larger for blindness in one eye than for any other complication. There may also be differences between studies in the definition of blindness. In the UKPDS, blindness was defined as LogMAR>1 in one eye, which may have very little impact on quality of life if vision in the

remaining eye is maintained. Disease progression in the worst eye may be the most clinically relevant measure, but visual acuity in the best eye may have more bearing on quality of life.

Finally, the current work employed linear regression methods to examine the impact of non-response and patient heterogeneity on estimates of quality of life change between health states. Linear models have well developed estimators for dealing with these issues, particularly in a panel data context. Nevertheless, it should be acknowledged that quality of life data have properties that may violate the assumptions underlying linear regression methods: for example, they are bounded above and below, and in the case of the EQ-5D have a conspicuous "ceiling effect" manifested in a significant proportion of the sample indicating no problems in any of the five domains (mobility, self-care, usual activity, pain/discomfort, anxiety/depression). Examining the impact of these data characteristics within a panel estimation framework would be an important future contribution, although it should be noted that that estimators that exploit within patient variation become increasingly intractable in the context of non-linearity. In the current study my primary concern has been not the distributional form, which mainly affects efficiency, but rather consistent estimation of causal effects in light of the potential bias arising from non-response and patient heterogeneity.

Appendix

A2. 1: Mean values for a meta-analysis of eight health states

Complication	No. of studies	Utility value range	Utility (95% CI)	Mean number of patients (range)
General diabetes	45	0.53–0.88	0.76 (0.75–0.77)	776 (22–7,348)
No complications	7	0.74–0.88	0.81 (0.78–0.84)	578.25 (40–2,636)
Myocardial infarction	5	0.68–0.77	0.75 (0.73–0.78)	127.7 (37–200)
Stroke	5	0.31–0.79	0.59 (0.41–0.77)	284.3 (34–701)
Ulcer	5	0.20–0.84	0.60 (0.48–0.72)	164.4 (19–701)
Amputation	5	0.34–0.740	0.56 (0.46–0.66)	205.1 (4–701)
Blindness	6	0.34–0.77	0.53 (0.40–0.66)	284.13 (3–701)
ESRD	4	0.33–0.81	0.48 (0.25–0.71)	184.6 (9–701)

Source: (Lung, Hayes et al. 2011)

A2. 2: A copy of the EuroQol Quality of Life questionnaire administered

UKPDS EuroQol Quality of Life questionnaire

Your health today

The following questions are about your general health. Please indicate which statement describes your own health state ***today***. Please ask a member of staff if you have any questions about this form.

Card no: 9 9 W Patient no: 0 0 0 0 1 H Visit no: 1 8 5 Date: 2 3 / 1 0 / 9 6

Please write one number in each section

Mobility	I have no problems in walking about	1	
	I have some problems in walking about	2	1
	I am confined to bed	3	
Self-care	I have no problems with self-care	1	
	I have some problems with washing or dressing myself	2	1
	I am unable to wash and dress myself	3	
Usual activities	I have no problems in performing my usual activities (e.g. work, study, housework, leisure activity)	1	
	I have some problems in performing my usual activities	2	1
	I am unable to perform my usual activities	3	
Pain/discomfort	I have no pain or discomfort	1	
	I have moderate pain or discomfort	2	1
	I have extreme pain or discomfort	3	
Anxiety/depression	I am not anxious or depressed	1	
	I am moderately anxious or depressed	2	1
	I am extremely anxious or depressed	3	

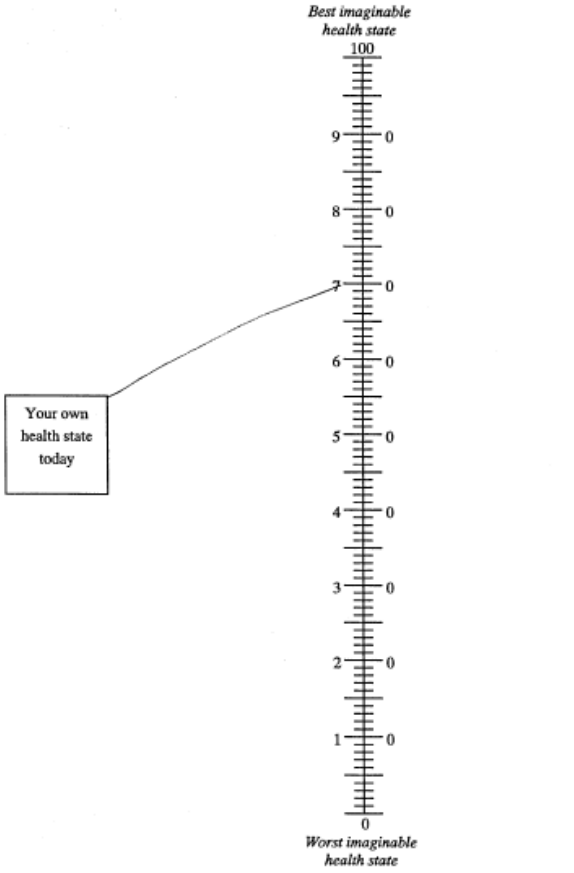
Please continue on the next page ☞

For office use only *Result:* 1 1 1 1 1 *Health state:* 7 0

COPY

To help people say how good or bad a health state is, we have drawn a scale (rather like a thermometer) on which the best state you can imagine is marked by 100, and the worst state is marked by 0.

We would like you to indicate on this scale how good or bad is your own health today, in your opinion. Please do this by drawing a line from the box below to whichever point on the scale indicates how good or bad your health state is.



A2. 3: Utility weights according to Dolan et al. (1996)

Attribute	Level	Decrement
constant		-0.081
	1	0
	2	-0.069
Mobility	**3**	-0.314
	1	0
	2	-0.104
Self-care	**3**	-0.214
	1	0
	2	-0.036
Usual Activities	**3**	-0.094
	1	0
	2	-0.123
Pain / Discomfort	**3**	-0.386
	1	0
	2	-0.071
Anxiety / Depression	**3**	-0.236
N3	If level 3	-0.269

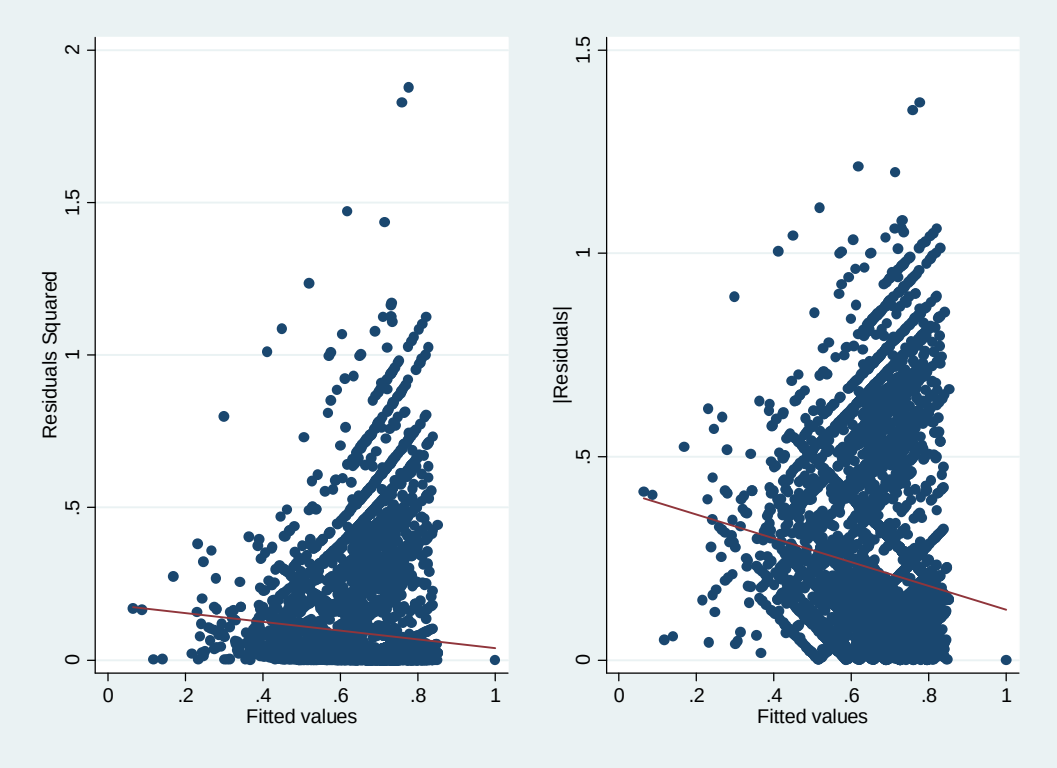
For example, assuming a patient has the following valuation: **11223**, then his answers would be mapped into a unique tariff, where utility weights are assigned using the social tariff developed by Dolan et al. (1996) as in the above table:
$tariff = 1 - (0.081(\text{constant}) + 0.036(\text{usual activity}) + 0.123(\text{pain}) + 0.236(\text{depression}) + 0.269) = 0.491$

Notice the presence of an extra negative coefficient if any given answer is equal to 3 (N3). This is because in the Dolan study respondents appeared to be more prepared to sacrifice life expectancy for states that include extreme problems with any of the dimensions. Level 2, which involves some problems, on the dimensions appears to be much more tolerable. For example, state 22222 has a median valuation that is 0.13 higher than 11113 and 0.25 higher than 11131. This results in most states that include level 3 on two or more dimensions having values that imply they are, on average, perceived to be worse than death.

A2. 4: Quality of life questionnaire response rate (2003-2007) across the 23 UKPDS centres

	questionnaire 2		questionnaire 3		questionnaire 4		questionnaire 5		questionnaire 6		questionnaire 7		Total		
Centre	issued	returned	issued	returned	issued	returned	issued	returned	issued	returned	issued	returned	**issued**	**returned**	**Rate**
Aberdeen	94	72	97	78	91	68	86	68	67	49	73	52	508	387	76%
London	108	67	102	73	98	67	97	62	72	53	89	61	566	383	68%
Belfast City	62	37	60	43	57	36	54	33	43	25	51	31	327	205	63%
Belfast Victoria	126	79	123	90	120	77	117	71	87	57	106	64	679	438	65%
Birmingham	122	90	113	84	109	76	103	71	85	55	98	65	630	441	70%
Carshalton	103	85	101	83	98	82	96	79	69	59	85	74	552	462	84%
Derby	79	47	78	54	78	52	75	39	51	23	69	38	430	253	59%
Dundee	115	97	114	99	111	97	107	89	83	67	99	86	629	535	85%
Hammersmith	84	51	82	43	78	45	74	39	55	28	67	37	440	243	55%
Ipswich	113	93	109	91	99	87	98	82	76	61	86	72	581	486	84%
Leicester	51	42	49	42	48	37	44	35	31	24	43	29	266	209	79%
Manchester	41	34	45	37	45	37	44	35	26	22	39	31	240	196	82%
Northampton	121	104	115	98	111	97	106	79	79	64	95	83	627	525	84%
Norwich	124	100	118	102	112	94	105	88	80	62	99	75	638	521	82%
Oxford	180	144	174	150	166	143	164	134	128	94	151	121	963	786	82%
Peterborough	41	39	40	39	39	36	37	33	28	25	33	31	218	203	93%
Salford	103	60	103	64	102	58	99	53	83	51	99	57	589	343	58%
St. George's	244	161	234	153	225	142	219	130	164	93	199	113	1285	792	62%
Stevenage	71	63	69	61	67	61	65	58	50	46	61	49	383	338	88%
Stoke-on-Trent	105	81	100	88	97	81	92	73	72	59	81	64	547	446	82%
Torbay	63	39	61	46	60	43	60	43	50	38	56	41	350	250	71%
Whittington	92	76	88	77	85	74	83	69	66	50	77	64	491	410	84%
TOTAL	**2150**	**1585**	**2087**	**1618**	**2011**	**1516**	**1942**	**1394**	**1479**	**1055**	**1779**	**1274**	**11448**	**8442**	**74%**

A2. 5: Heteroskedasticity: sample residuals versus fitted values plots



Herein the residuals plotted against the fitted values. As the tariff increases, there is a trend towards lower absolute residuals.

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity. Ho: Constant variance. Variables: fitted values of tariff
F(1 , 11612) = 74.74; Prob > F = 0.0000.

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity. Ho: Constant variance. Variables: age_now qother MI MI_prior past_ihd past_stroke past_hf past_amp past_blind. $\chi 2$ (9) = 95.72;
Prob > $\chi 2$ = 0.0000

A2. 6: Differences between singletons and multiple responders at first questionnaires (1996/7)

		average			**singletons**			**responders to multiple questionnaires**			**Difference (Std. Err.)**
	Variables	mean	sd	N	mean	sd	N	mean	sd	N	***P-value<0.05**
QoL	tariff	0.76	0.27	3243	0.73	0.29	1410	0.79	0.25	1833	-0.069 (0.009)*
	VAS	73.55	19.05	3192	70.93	19.69	1387	75.56	18.30	1805	-4.625 (0.675)*
demographics	current age	62.29	9.00	3243	63.45	9.49	1410	61.39	8.51	1833	2.058 (0.317)*
	duration	10.60	2.84	3243	10.74	2.89	1410	10.50	2.80	1833	0.239 (0.101)*
	male	58.7%	0.49	3243	59.0%	0.49	1410	58.5%	0.49	1833	0.005 (0.017)
race	White Caucasian	81.5%	0.389	3243	82.1%	0.384	1410	81.0%	0.392	1833	0.010 (0.014)
	Asian Indian	7.3%	0.261	3243	7.5%	0.264	1410	7.2%	0.259	1833	0.003 (0.009)
	Afro-Caribbean	10.7%	0.309	3243	10.1%	0.302	1410	11.1%	0.314	1833	-0.009 (0.011)
	Asian Chinese	0.2%	0.043	3243	0.2%	0.046	1410	0.2%	0.040	1833	0.000 (0.002)
	Other	0.3%	0.058	3243	0.1%	0.027	1410	0.5%	0.074	1833	-0.005 (0.002)*
complications	MI	0.064	0.276	3243	0.078	0.305	1410	0.053	0.250	1833	0.025 (0.010)*
	IHD	0.054	0.227	3243	0.063	0.243	1410	0.047	0.214	1833	0.016 (0.008)*
	Stroke	0.019	0.145	3243	0.035	0.191	1410	0.007	0.093	1833	0.028 (0.005)*
	HF	0.016	0.124	3243	0.027	0.162	1410	0.007	0.084	1833	0.020 (0.004)*
	Amputation	0.007	0.096	3243	0.011	0.124	1410	0.003	0.066	1833	0.008 (0.003)*
	Renal Failure	0.004	0.061	3243	0.005	0.070	1410	0.003	0.052	1833	0.002 (0.002)
	Blindness in 1 eye	0.029	0.176	3243	0.036	0.201	1410	0.024	0.153	1833	0.012 (0.006)
	Cataract Extraction	0.060	0.302	3243	0.072	0.340	1410	0.051	0.269	1833	0.022 (0.011)*
	Retinal Photocoagulation	0.073	0.263	3243	0.089	0.287	1410	0.061	0.243	1833	0.028 (0.009)*
	Vitreous haemorrhage	0.008	0.101	3243	0.010	0.119	1410	0.007	0.084	1833	0.003 (0.004)
socioeconomic status	professional	4.4%	0.205	3149	4.6%	0.210	1364	4.3%	0.202	1785	0.004 (0.007)
	managerial	22.4%	0.417	3149	20.2%	0.402	1364	24.1%	0.428	1785	-0.039 (0.015)*
	skilled worker	43.3%	0.496	3149	42.4%	0.494	1364	43.9%	0.496	1785	-0.014 (0.018)
	semi-skilled	23.4%	0.423	3149	24.7%	0.431	1364	22.4%	0.417	1785	0.023 (0.015)
	unskilled	6.5%	0.247	3149	8.0%	0.271	1364	5.4%	0.226	1785	0.026 (0.009)*
biomarkers	hba1c	7.754	1.246	3243	7.790	1.261	1410	7.726844	1.235	1833	0.064 (0.044)
	sbp	137.206	14.804	3243	138.201	15.528	1410	136.4404	14.178	1833	1.760 (0.524)*
	ldl	3.425	0.805	3243	3.454	0.808	1410	3.403618	0.803	1833	0.050 (0.029)
	hdl	1.099	0.250	3243	1.090	0.249	1410	1.106118	0.250	1833	-0.016 (0.009)
	alb	70.886	188.427	3243	85.501	215.130	1410	59.6435	164.150	1833	25.857 (6.660)*
behavioural variables	BMI at baseline	28.77	5.51	3233	28.99	5.73	1404	28.60	5.33	1829	0.393 (0.195)*
	sedentary	9.8%	0.297	3243	12.1%	0.327	1410	8.0%	0.271	1833	0.042 (0.010)*
	moderately active	50.5%	0.500	3243	53.3%	0.499	1410	48.4%	0.500	1833	0.049 (0.018)*
	active	37.8%	0.485	3243	32.6%	0.469	1410	41.8%	0.493	1833	-0.092 (0.017)*
	fit	1.9%	0.136	3243	2.0%	0.140	1410	1.8%	0.133	1833	0.002 (0.005)
	none	30.2%	0.459	3243	32.0%	0.467	1410	28.9%	0.453	1833	0.031 (0.016)
	occasional	48.7%	0.500	3243	46.0%	0.499	1410	50.8%	0.500	1833	-0.049 (0.018)*
	regular	14.2%	0.349	3243	14.8%	0.355	1410	13.8%	0.345	1833	0.009 (0.012)
	dependent	6.8%	0.253	3243	7.3%	0.260	1410	6.5%	0.246	1833	0.008 (0.009)

A2. 7: Exploring basic patients' characteristics and attrition

	pts with more than 2 questionnaires			**pts with exactly 2**			**difference**	
	mean	***SD***	***total pts***	***mean***	***SD***	***total pts***	**btw >2 and 2**	**P>|t|**
tariff	0.681	0.251	1746	0.626	0.291	209	0.055	0.003
current age	68.107	8.551	1746	67.557	8.210	209	0.550	0.378
duration	17.524	2.778	1746	15.219	3.294	209	2.305	0.000
	pts with more than3 questionnaires			**pts with exactly 3**			**btw >3 and 3**	
tariff	0.681	0.251	1746	0.630	0.274	219	0.051	0.001
current age	68.107	8.551	1746	67.193	9.124	219	0.913	0.091
duration	17.524	2.778	1746	16.266	3.045	219	1.257	0.000
	pts with more than 4 questionnaires			**pts with exactly 4**			**btw >4 and 4**	
tariff	0.701	0.239	1324	0.600	0.273	203	0.101	0.000
current age	68.321	8.434	1324	67.691	8.636	203	0.630	0.323
duration	17.839	2.617	1324	16.826	2.990	203	1.012	0.000
	pts with more than 5 questionnaires			**pts with exactly 5**			**btw >5 and 5**	
tariff	0.714592	0.233132	1023	0.65687	0.25437	301	0.058	0.000
current age	68.53104	8.203372	1023	67.6087	9.15471	301	0.922	0.095
duration	17.83406	2.472206	1023	17.8537	3.06356	301	-0.020	0.909
	pts with 7 questionnaires			**pts with exactly 6**			**btw 7 and 6**	
tariff	0.724128	0.238001	535	0.70414	0.22746	488	0.020	0.171
current age	68.46549	8.297298	535	68.6029	8.10705	488	-0.137	0.789
duration	17.66616	2.307191	535	18.0181	2.63139	488	-0.352	0.023

A2. 8: Results of the Selection part of the Sample Selection Model (Heckman 1979) and Complete Results of the Outcome part of the Sample Selection Model, including location

Variables	Selection Coeff.	robust SE	Outcome tariff	robust SE
Constant	2.544**	(0.162)	0.866**	(0.042)
Age at Q1	-0.015**	(0.002)	-0.001*	(0.001)
Q2	-1.879**	(0.053)	-0.063**	(0.011)
Q3	-1.867**	(0.057)	-0.073**	(0.010)
Q4	-1.987**	(0.064)	-0.086**	(0.011)
Q5	-2.099**	(0.074)	-0.089**	(0.012)
Q6	-2.176**	(0.086)	-0.107**	(0.013)
Q7	-2.253**	(0.096)	-0.099**	(0.013)
Male(=1)	-0.012	(0.037)	0.073**	(0.010)
Asian	-0.224**	(0.068)	-0.032	(0.019)
Manual worker+	-0.087	(0.046)	-0.052**	(0.012)
Unskilled worker+	-0.237**	(0.049)	-0.064**	(0.013)
MI (year before)	0.619**	(0.139)	-0.112**	(0.038)
MI (prior history)	-0.024	(0.061)	-0.037*	(0.018)
IHD	0.100	(0.059)	-0.072**	(0.017)
Stroke	-0.296**	(0.081)	-0.180**	(0.032)
Heart Failure	-0.407**	(0.089)	-0.175**	(0.034)
Amputation	-0.476**	(0.124)	-0.182**	(0.042)
Blindness in 1 eye	0.018	(0.076)	-0.034	(0.023)
Aberdeen	-0.124	(0.117)	-0.001	(0.029)
Belfast City	0.307*	(0.137)	-0.054	(0.035)
Belfast Victoria	0.188	(0.108)	-0.079**	(0.030)
Birmingham	0.149	(0.095)	-0.022	(0.029)
Carshalton	0.294**	(0.109)	0.027	(0.028)
Derby	0.252	(0.148)	-0.077	(0.040)
Dundee	0.604**	(0.129)	0.035	(0.030)
Hammersmith	0.249	(0.130)	-0.031	(0.033)
Ipswich	0.379**	(0.111)	0.023	(0.030)
Leicester	0.288	(0.176)	0.010	(0.039)
Manchester	0.106	(0.191)	-0.001	(0.040)
Northampton	0.800**	(0.147)	0.006	(0.030)
Norwich	0.545**	(0.114)	0.032	(0.027)
Oxford	0.424**	(0.101)	0.009	(0.026)
Peterborough	0.517*	(0.218)	0.067	(0.038)
Salford	0.203	(0.127)	-0.055	(0.038)
Scarborough	1.151	(0.848)	0.018	(0.034)
St. George's	0.433**	(0.097)	0.001	(0.027)
Stevenage	-0.098	(0.144)	0.045	(0.034)
Stoke-on-Trent	0.519**	(0.143)	-0.074*	(0.035)
Torbay	0.186	(0.137)	0.038	(0.032)
Whittington	0.259*	(0.118)	-0.003	(0.034)
Aberdeen II	0.000	(0.000)	0.000	(0.000)
Q*Aberdeen	0.129**	(0.029)		
Q*Belfast City	-0.024	(0.028)		
Q*Belfast Victoria	0.016	(0.023)		
Q*Birmingham	-0.003	(0.022)		
Q*Carshalton	0.035	(0.021)		
Q*Derby	-0.000	(0.030)		

Q*Dundee	0.029	(0.024)
Q*Hammersmith	-0.005	(0.027)
Q*Ipswich	0.011	(0.021)
Q*Leicester	0.020	(0.037)
Q*Manchester	0.090*	(0.043)
Q*Northampton	0.011	(0.028)
Q*Norwich	0.002	(0.022)
Q*Oxford	0.033	(0.021)
Q*Peterborough	0.027	(0.042)
Q*Salford	0.067**	(0.025)
Q*Scarborough	-1.099*	(0.546)
Q*St. George's	-0.000	(0.020)
Q*Stevenage	0.160**	(0.035)
Q*Stoke-on-Trent	0.040	(0.028)
Q*Torbay	0.065*	(0.025)
Q*Whittington	0.032	(0.024)
Q*Aberdeen II	0.000	(0.000)
Observations	22274	
Censored	10946	
Uncensored	11328	

+comparator: professional and managerial

** p<0.01, * p<0.05

Wald test of indep. eqns. (rho = 0): $\chi2(1) = 1.18$ Prob > $\chi2 = 0.2775$

lambda	-0.011	0.010

Joint test that all the interactions between locations and the time trend (Q*location) are equal to zero: F(22, 28821) = 4.54. Prob > F =0.000 supporting the use of the proposed exclusion restriction.

A2. 9: Test of OLS with Selection and FE with Selection as in Wooldridge 1995

(Page 130, Test 4.2)

Variables	Coef.	Robust SE
Constant	0.839**	(0.047)
Age at Q1	-0.001	(0.001)
Q2	-0.112**	(0.033)
Q3	-0.067	(0.034)
Q4	-0.069*	(0.035)
Q5	-0.072	(0.040)
Q6	-0.078	(0.044)
Q7	-0.041	(0.051)
Male(=1)	0.075**	(0.012)
Asian	-0.049**	(0.018)
MI (year before)	-0.108*	(0.043)
MI (prior history)	-0.013	(0.030)
IHD	-0.017	(0.026)
Stroke	-0.168**	(0.041)
Heart Failure	-0.129**	(0.041)
Amputation	-0.131**	(0.046)
Blindness in 1 eye	-0.017	(0.034)
MIyr1	0.074	(0.116)
MIyr2	-0.155*	(0.078)
MIyr3	-0.252	(0.156)
MIyr4	0.028	(0.068)
MIyr5	-0.134	(0.152)
MIyr6	-0.019	(0.099)
MIyr7	0.112	(0.081)
MI_prioryr1	-0.005	(0.046)
MI_prioryr2	-0.232	(0.120)
MI_prioryr3	-0.046	(0.164)
MI_prioryr4	0.343*	(0.159)
MI_prioryr5	-0.253	(0.160)
MI_prioryr6	0.048	(0.178)
MI_prioryr7	0.113	(0.107)
past_ihdyr1	-0.022	(0.039)
past_ihdyr2	0.106	(0.081)
past_ihdyr3	-0.233**	(0.088)
past_ihdyr4	0.185**	(0.067)
past_ihdyr5	-0.051	(0.070)
past_ihdyr6	-0.009	(0.070)
past_ihdyr7	-0.048	(0.056)
past_stroke1	-0.069	(0.066)
past_stroke2	0.126	(0.075)
past_stroke3	-0.093	(0.107)
past_stroke4	0.147	(0.140)
past_stroke5	-0.100	(0.156)
past_stroke6	-0.190	(0.123)
past_stroke7	0.127*	(0.056)
past_hfyr1	0.021	(0.076)
past_hfyr2	-0.028	(0.077)
past_hfyr3	-0.050	(0.077)
past_hfyr4	0.111	(0.107)
past_hfyr5	0.073	(0.165)
past_hfyr6	-0.144	(0.156)
past_hfyr7	-0.024	(0.090)

past_ampyr1	-0.091	(0.099)
past_ampyr2	0.047	(0.118)
past_ampyr3	0.213	(0.128)
past_ampyr4	-0.087	(0.107)
past_ampyr5	-0.158	(0.105)
past_ampyr6	0.168*	(0.065)
past_ampyr7	-0.171**	(0.031)
past_blind1	0.014	(0.054)
past_blind2	-0.053	(0.099)
past_blind3	0.049	(0.163)
past_blind4	-0.016	(0.149)
past_blind5	-0.055	(0.115)
past_blind6	0.170	(0.101)
past_blind7	-0.144**	(0.042)
invmills1	-0.019	(0.035)
invmills2	0.038	(0.034)
invmills3	-0.028	(0.036)
invmills4	-0.035	(0.036)
invmills5	-0.038	(0.038)
invmills6	-0.045	(0.040)
invmills7	-0.075	(0.046)

** p<0.01, * p<0.05

Wald test for joint significance of the time varying variables per questionnaire round F(49, 2662)=3.03, Prob > F = 0.0000.

A2. 10: Results of the Selection part of the Panel Sample Selection Model (Wooldridge 1995)

	Q1	Q2	Q3	Q4	Q5	Q6	Q7
VARIABLES	selected	selected	selected	selected	selected	selected	selected
MIyr1	-0.140	-0.680**	-0.716**	-0.705**	-0.619*	-0.248	-0.192
	(0.327)	(0.331)	(0.332)	(0.347)	(0.339)	(0.316)	(0.329)
MIyr2	4.816	1.339*	0.841	0.123	0.781	0.251	-0.178
	(311.468)	(0.690)	(0.648)	(0.662)	(0.628)	(0.669)	(0.722)
MIyr3	-1.092	-5.756	-0.285	-0.818	-0.779	-4.810	-0.594
	(0.684)	(373.325)	(0.699)	(0.747)	(0.804)	(150.411)	(0.753)
MIyr4	3.908	-0.861	0.410	-0.816	0.233	-0.373	-4.890
	(192.723)	(0.769)	(0.914)	(1.036)	(0.848)	(0.791)	(218.988)
MIyr5	4.428	5.748	0.504	5.631	0.167	1.034	1.119
	(317.805)	(293.237)	(0.636)	(227.369)	(0.671)	(0.775)	(0.749)
MIyr6	-0.481	0.579	0.285	-1.013	0.802	0.181	0.211
	(1.007)	(0.812)	(0.755)	(0.795)	(0.773)	(0.723)	(0.715)
MIyr7	-0.155	0.029	-0.347	0.560	0.275	0.081	-0.218
	(0.670)	(0.543)	(0.523)	(0.555)	(0.549)	(0.523)	(0.576)
MI_prioryr1	-0.131	-0.451**	-0.397**	-0.222	-0.151	-0.355*	-0.102
	(0.207)	(0.178)	(0.177)	(0.179)	(0.178)	(0.181)	(0.184)
MI_prioryr2	4.560	1.073	0.649	-0.243	0.082	-0.340	-0.785
	(311.469)	(0.795)	(0.757)	(0.774)	(0.743)	(0.775)	(0.820)
MI_prioryr3	-5.209	-6.604	-0.720	-0.542	-0.677	-4.989	0.179
	(311.469)	(373.325)	(1.015)	(1.058)	(1.051)	(150.412)	(1.109)
MI_prioryr4	4.230	4.603	0.004	-0.376	0.949	4.715	-4.597
	(192.725)	(373.325)	(1.214)	(1.334)	(1.196)	(150.412)	(218.990)
MI_prioryr5	0.442	5.958	-0.309	5.214	-1.251	1.500	5.902
	(371.675)	(293.238)	(1.160)	(227.371)	(1.123)	(1.136)	(218.990)

MI_prioryr6	-8.424	-4.472	0.079	-5.105	1.048	-1.275	-0.628
	(433.724)	(293.239)	(1.087)	(227.371)	(1.152)	(1.175)	(1.130)
MI_prioryr7	4.612	-0.197	0.576	1.247	0.092	0.643	-0.012
	(295.155)	(0.900)	(0.848)	(0.911)	(0.854)	(0.809)	(0.780)
past_ihdyr1	-0.393**	-0.637***	-0.482***	-0.409***	-0.379**	-0.401***	-0.273*
	(0.191)	(0.156)	(0.154)	(0.154)	(0.154)	(0.155)	(0.156)
past_ihdyr2	-0.405	-0.222	-0.502	-0.615	-0.468	-0.526	-0.381
	(0.596)	(0.394)	(0.416)	(0.416)	(0.403)	(0.402)	(0.388)
past_ihdyr3	0.503	0.221	0.455	0.194	0.468	0.326	0.558
	(0.744)	(0.539)	(0.558)	(0.563)	(0.546)	(0.543)	(0.548)
past_ihdyr4	-0.054	-0.481	-0.074	0.437	-0.058	0.407	-0.411
	(0.802)	(0.634)	(0.599)	(0.592)	(0.581)	(0.577)	(0.594)
past_ihdyr5	-0.269	-0.128	-1.175*	-1.083*	-1.538**	-1.120*	-0.751
	(0.816)	(0.626)	(0.675)	(0.602)	(0.667)	(0.575)	(0.570)
past_ihdyr6	-3.446	0.881	-3.587	-4.177	-3.662	0.779	0.823
	(541.970)	(0.996)	(189.118)	(368.690)	(242.233)	(1.004)	(0.926)
past_ihdyr7	4.228	0.293	5.365	5.680	5.636	0.556	0.494
	(541.969)	(0.923)	(189.118)	(368.690)	(242.233)	(0.929)	(0.849)
past_strokeyr1	-0.688***	-0.683***	-0.708***	-0.620***	-0.983***	-0.624**	-0.709***
	(0.208)	(0.221)	(0.223)	(0.228)	(0.271)	(0.249)	(0.274)
past_strokeyr2	-7.324	0.542	-0.042	-0.838	-0.378	-0.133	-0.302
	(278.386)	(0.508)	(0.487)	(0.515)	(0.484)	(0.489)	(0.502)
past_strokeyr3	7.814	-0.686	-0.191	0.809	0.544	0.051	0.814
	(278.386)	(0.674)	(0.658)	(0.670)	(0.681)	(0.683)	(0.774)
past_strokeyr4	-0.637	-0.974	-1.160*	-0.934	-0.394	-0.108	-1.167*
	(0.686)	(0.624)	(0.647)	(0.591)	(0.609)	(0.612)	(0.698)
past_strokeyr5	-4.510	0.073	0.315	0.313	-1.337*	-0.420	0.461
	(279.683)	(0.746)	(0.758)	(0.655)	(0.735)	(0.630)	(0.646)
past_strokeyr6	-0.259	0.875	-4.032	0.316	-3.800	5.071	-0.393
	(594.479)	(1.108)	(190.221)	(1.053)	(244.117)	(226.075)	(1.051)
past_strokeyr7	5.033	0.277	5.154	0.306	5.286	-4.563	0.381
	(524.579)	(0.918)	(190.220)	(0.908)	(244.116)	(226.074)	(0.896)
past_hfyr1	0.003	-0.897***	-0.543**	-0.623**	-0.461*	-0.707**	-0.594*
	(0.267)	(0.270)	(0.259)	(0.273)	(0.270)	(0.318)	(0.304)
past_hfyr2	-0.038	0.325	0.195	0.424	0.321	0.268	0.065
	(0.644)	(0.490)	(0.492)	(0.533)	(0.525)	(0.542)	(0.538)
past_hfyr3	-4.734	-1.564**	-0.922	-0.613	-0.245	-1.119	-0.289
	(284.458)	(0.772)	(0.681)	(0.699)	(0.707)	(0.698)	(0.713)
past_hfyr4	0.077	-4.225	-0.325	-0.875	-0.625	-0.475	-0.569
	(423.549)	(296.458)	(0.813)	(0.788)	(0.763)	(0.819)	(0.767)
past_hfyr5	4.612	4.501	-0.134	0.413	-0.823	-0.292	-0.161
	(313.812)	(296.458)	(0.905)	(0.862)	(0.862)	(0.932)	(0.824)
past_hfyr6	-4.131	-0.009	0.144	-0.556	0.090	1.500	5.401
	(380.790)	(0.963)	(0.947)	(0.923)	(0.952)	(0.945)	(204.806)
past_hfyr7	4.387	1.125	1.034	1.143	1.234*	-0.104	-4.632
	(380.789)	(0.706)	(0.701)	(0.706)	(0.705)	(0.686)	(204.805)
past_ampyr1	-0.637*	-0.179	-0.784*	-0.650	-0.498	-0.568	
	(0.354)	(0.453)	(0.456)	(0.465)	(0.425)	(0.466)	
past_ampyr2	0.011	-1.062**	-1.842***	-0.821*	-0.073	-0.287	-0.333
	(0.546)	(0.503)	(0.610)	(0.493)	(0.495)	(0.499)	(0.501)
past_ampyr3	-0.525	-0.508	1.113	-0.095	0.543	0.617	0.443
	(0.815)	(0.669)	(0.745)	(0.661)	(0.747)	(0.757)	(0.759)
past_ampyr4	-4.340	1.346*	-0.189	-0.706	-1.404*	-0.365	-0.300
	(324.817)	(0.792)	(0.747)	(0.816)	(0.827)	(0.860)	(0.858)
past_ampyr5	0.231	-1.140	0.350	0.975	0.191	-1.148	-0.426

	(409.303)	(0.802)	(0.740)	(0.792)	(0.737)	(0.785)	(0.754)
past_ampyr6	-0.081	5.726	-4.717	-5.108	-4.685	-4.295	-5.125
	(807.399)	(778.120)	(269.196)	(523.609)	(345.307)	(263.135)	(341.203)
past_ampyr7	4.664	-5.089	4.949	5.363	5.225	5.168	5.339
	(768.030)	(778.120)	(269.195)	(523.609)	(345.307)	(263.134)	(341.203)
past_blindyr1	-0.380	-0.285	-0.255	-0.416**	-0.112	-0.238	-0.248
	(0.244)	(0.200)	(0.198)	(0.202)	(0.206)	(0.211)	(0.217)
past_blindyr2	-3.795	0.163	0.192	0.707	-0.263	-1.013	-0.275
	(337.415)	(0.656)	(0.670)	(0.666)	(0.637)	(0.737)	(0.586)
past_blindyr3	-0.751	-0.757	-0.116	-1.326	-0.670	-4.199	-0.816
	(499.842)	(0.942)	(0.910)	(0.943)	(0.923)	(131.393)	(0.888)
past_blindyr4	0.177	0.489	-0.649	0.001	-0.655	4.151	-0.637
	(499.308)	(0.887)	(0.841)	(0.884)	(0.923)	(131.393)	(0.925)
past_blindyr5	4.718	-0.593	0.468	0.400	1.089	0.534	1.078
	(336.625)	(0.892)	(0.786)	(0.796)	(0.835)	(0.790)	(0.834)
past_blindyr6	-4.498	-5.310	-5.508	-5.594	-5.544	-5.302	-5.374
	(768.031)	(681.404)	(269.196)	(523.609)	(345.307)	(263.135)	(341.203)
past_blindyr7	4.626	6.302	5.891	6.141	6.131	5.954	6.047
	(768.030)	(681.404)	(269.195)	(523.609)	(345.307)	(263.134)	(341.203)
Male(=1)	0.080	0.060	0.020	0.052	0.046	0.068	0.044
	(0.051)	(0.048)	(0.047)	(0.048)	(0.048)	(0.049)	(0.049)
Asian	0.162*	-0.205**	-0.167*	-0.270***	-0.146*	-0.090	-0.118
	(0.096)	(0.088)	(0.087)	(0.089)	(0.088)	(0.089)	(0.090)
Age at Q1	0.005*	-0.008***	-0.011***	-0.010***	-0.015***	-0.013***	-0.015***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Aberdeen	-0.855***	0.179	0.194	0.254*	0.253*	0.332**	0.367**
	(0.151)	(0.148)	(0.145)	(0.148)	(0.147)	(0.150)	(0.152)
Belfast City	-0.418**	-0.009	0.070	0.105	-0.113	0.010	-0.024
	(0.171)	(0.169)	(0.164)	(0.168)	(0.173)	(0.175)	(0.180)
Belfast Victoria	-0.305**	0.194	0.185	0.218	0.034	0.216	0.144
	(0.148)	(0.138)	(0.135)	(0.138)	(0.139)	(0.141)	(0.145)
Birmingham	0.103	0.234*	0.069	0.152	0.012	0.112	0.198
	(0.151)	(0.133)	(0.132)	(0.136)	(0.136)	(0.139)	(0.141)
Carshalton	-0.125	0.337**	0.305**	0.448***	0.356**	0.363**	0.428***
	(0.157)	(0.141)	(0.139)	(0.141)	(0.141)	(0.144)	(0.146)
Derby	-0.113	0.194	0.260	0.353**	0.075	-0.148	0.086
	(0.193)	(0.169)	(0.167)	(0.169)	(0.172)	(0.181)	(0.179)
Dundee	0.177	0.691***	0.671***	0.760***	0.561***	0.658***	0.728***
	(0.175)	(0.144)	(0.143)	(0.145)	(0.144)	(0.147)	(0.149)
Hammersmith	-0.678***	-0.069	-0.190	0.025	-0.245	-0.168	-0.072
	(0.155)	(0.157)	(0.156)	(0.157)	(0.162)	(0.166)	(0.166)
Ipswich	0.196	0.476***	0.298**	0.362**	0.214	0.258*	0.404***
	(0.171)	(0.141)	(0.139)	(0.142)	(0.142)	(0.146)	(0.146)
Leicester	0.027	0.640***	0.696***	0.634***	0.586***	0.506**	0.464**
	(0.236)	(0.203)	(0.203)	(0.203)	(0.204)	(0.204)	(0.207)
Manchester	-0.670***	0.195	0.290	0.404**	0.187	0.262	0.316
	(0.192)	(0.191)	(0.187)	(0.189)	(0.191)	(0.194)	(0.196)
Northampton	0.014	0.827***	0.628***	0.769***	0.374**	0.576***	0.698***
	(0.176)	(0.152)	(0.149)	(0.151)	(0.150)	(0.152)	(0.154)
Norwich	-0.411***	0.434***	0.378***	0.394***	0.290**	0.237*	0.336**
	(0.148)	(0.138)	(0.136)	(0.139)	(0.139)	(0.143)	(0.145)
Oxford	-0.479***	0.468***	0.429***	0.492***	0.358***	0.299**	0.471***
	(0.134)	(0.125)	(0.123)	(0.126)	(0.125)	(0.129)	(0.130)
Peterborough	-0.721***	0.264	0.250	0.338*	0.060	0.188	0.336*
	(0.182)	(0.180)	(0.178)	(0.180)	(0.185)	(0.185)	(0.187)

Salford	-0.243	0.301*	0.285*	0.404**	0.165	0.386**	0.410**
	(0.175)	(0.161)	(0.158)	(0.161)	(0.161)	(0.162)	(0.165)
Scarborough	-0.446***	-1.791***	-1.888***	-1.721***			
	(0.169)	(0.397)	(0.398)	(0.397)			
St. George's	-0.136	0.437***	0.262**	0.381***	0.204	0.210	0.282**
	(0.139)	(0.125)	(0.124)	(0.127)	(0.127)	(0.131)	(0.132)
Stevenage	-1.012***	0.407***	0.402***	0.492***	0.319**	0.423***	0.348**
	(0.161)	(0.156)	(0.155)	(0.157)	(0.158)	(0.160)	(0.164)
Stoke-on-Trent	-0.350**	0.428***	0.498***	0.481***	0.395***	0.442***	0.455***
	(0.157)	(0.144)	(0.142)	(0.145)	(0.145)	(0.148)	(0.150)
Torbay	-0.518***	0.149	0.199	0.307*	0.164	0.347**	0.342**
	(0.173)	(0.167)	(0.163)	(0.166)	(0.165)	(0.167)	(0.169)
Whittington	-0.251*	0.264*	0.217	0.313**	0.181	0.210	0.314**
	(0.151)	(0.140)	(0.138)	(0.141)	(0.141)	(0.145)	(0.146)
Aberdeen II	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.597***	-0.215	0.108	-0.120	0.262	0.035	0.134
	(0.204)	(0.193)	(0.191)	(0.194)	(0.195)	(0.198)	(0.199)
	3355	3355	3355	3355	3254	3254	3231

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note that the questionnaire dummies in q11, q12, …q17, q21, q22, etc are always 1 or 0 for all patients by construction and so are collinear with the constant term and therefore dropped as controls.

A2. 11: Table 2.12 and 2.13 OLS with location

Dependent variable: Tariff	Table 2.12 (grouping complications) Coeff	SE	Table 2.13 (all complications) Coeff	SE
Constant	0.853**	-0.023	0.849**	-0.039
Age at Q1	-0.001*	-0.001	-0.001	-0.001
Q2	-0.071**	-0.008	-0.070**	-0.008
Q3	-0.081**	-0.008	-0.080**	-0.008
Q4	-0.097**	-0.008	-0.095**	-0.008
Q5	-0.099**	-0.009	-0.097**	-0.008
Q6	-0.116**	-0.01	-0.115**	-0.01
Q7	-0.109**	-0.009	-0.109**	-0.009
Complication (=1)	-0.106**	-0.012		
MI (year before)			-0.102**	-0.037
MI (prior history)			-0.038*	-0.018
IHD			-0.073**	-0.016
Stroke			-0.193**	-0.03
Heart Failure			-0.166**	-0.031
Amputation			-0.201**	-0.039
Blindness in 1 eye			-0.038	-0.022
Male(=1)	0.074**	-0.01	0.074**	-0.01
Asian	-0.028	-0.019	-0.033	-0.019
Manual worker	-0.054**	-0.012	-0.053**	-0.012
Unskilled worker	-0.068**	-0.014	-0.066**	-0.013
Aberdeen	-0.051	(0.035)	0.000	(0.000)
Belfast City	-0.075*	(0.031)	-0.055*	(0.022)
Belfast Victoria	-0.015	(0.029)	-0.078**	(0.018)
Birmingham	0.029	(0.028)	-0.022	(0.018)
Carshalton	-0.074	(0.040)	0.026	(0.018)
Derby	0.040	(0.029)	-0.076**	(0.021)
Dundee	-0.024	(0.032)	0.037*	(0.017)
Hammersmith	0.029	(0.030)	-0.029	(0.021)
Ipswich	0.011	(0.041)	0.024	(0.018)
Leicester	0.003	(0.040)	0.009	(0.022)
Manchester	0.011	(0.031)	-0.002	(0.023)
Northampton	0.041	(0.027)	0.012	(0.018)
Norwich	0.013	(0.027)	0.034	(0.018)
Oxford	0.068	(0.039)	0.011	(0.016)
Peterborough	-0.047	(0.036)	0.069**	(0.023)
Salford	0.015	(0.035)	-0.052**	(0.020)
Scarborough	0.005	(0.026)	0.018	(0.034)
St. George's	0.051	(0.034)	0.004	(0.016)
Stevenage	-0.063	(0.037)	0.046*	(0.020)
Stoke-on-Trent	0.045	(0.033)	-0.071**	(0.019)
Torbay	-0.002	(0.034)	0.043*	(0.021)
Whittington	-0.002	(0.029)	-0.003	(0.018)
Aberdeen II	0.000	(0.000)	0.001	(0.018)

** p<0.01, * p<0.05

A2. 12: Robust Hausman test

Combining all complications (including all controls):

xi: xtreg tariff m_age_q1 q2 q3 q4 q5 q6 q7 complications last_male indian i.socclass i.cities, re vce(cluster ukno)
xtoverid

Test of overidentifying restrictions: fixed vs random effects
Cross-section time-series model: xtreg re robust cluster(patient)
Sargan-Hansen statistic= 80.572 Chi-sq(7) P-value = 0.0000

Combining all complications (no controls):

xtreg tariff m_age_q1 q2 q3 q4 q5 q6 q7 complications, re vce(cluster ukno)
xtoverid

Test of overidentifying restrictions: fixed vs random effects
Cross-section time-series model: xtreg re robust cluster(patient)
Sargan-Hansen statistic= 90.444 Chi-sq(7) P-value = 0.0000

All complications (including all controls):

global past "MI MI_prior past_ihd past_stroke past_hf past_amp past_blind"
xi:xtreg tariff m_age_q1 q2 q3 q4 q5 q6 q7 $past last_male indian i.socclass i.cities, re vce(cluster ukno)
xtoverid

Test of overidentifying restrictions: fixed vs random effects
Cross-section time-series model: xtreg re robust cluster(patient)
Sargan-Hansen statistic= 76.111 Chi-sq(13) P-value = 0.0000

All complications (no controls):

Test of overidentifying restrictions: fixed vs random effects
Cross-section time-series model: xtreg re robust cluster(ukno)
Sargan-Hansen statistic= 87.510 Chi-sq(13) P-value = 0.0000

A2. 13: Results using Visual Analogue Scale

FE		
Dependent variable: VAS	Coeff	SE
Constant	0.938**	(0.059)
Q2	-0.114**	(0.007)
Q3	-0.123**	(0.007)
Q4	-0.143**	(0.007)
Q5	-0.149**	(0.008)
Q6	-0.172**	(0.009)
Q7	-0.167**	(0.008)
MI (year before)	-0.098**	(0.027)
MI (prior history)	-0.018	(0.016)
Ischaemic heart disease	0.006	(0.014)
Stroke	-0.096**	(0.020)
Heart Failure	-0.024	(0.024)
Amputation	-0.079**	(0.026)
Blindness in 1 eye	-0.006	(0.023)
Observations	11450	
Number of patients	3357	
R-squared	0.088	

** p<0.01, * p<0.05

The VAS is rescaled from a 0 to 100 range to a 0 to 1 range in order to facilitate comparisons with the EQ-5D.
The impact of complications on the VAS score was smaller than on the tariff. Clarke et al. (2002) found that the rankings were maintained whether one uses the EQ-5D or the VAS. With additional rounds of data I find instead that rankings of the impact of complication on VAS and EQ-5D are not equivalent: amputations are given higher weights by the utility algorithm, whereas myocardial infarctions are ranked higher by patients on the VAS. This, as was also argued by Bagust et al. (2005), may reflect the high weighting attached to loss of mobility and severe pain in the time-trade-off algorithm.

References

Bagust, A. and S. Beale (2005). "Modelling EuroQol health-related utility values for diabetic complications from CODE-2 data." Health Economics 14(3): 217-230.

Bharmal, M. and J. Thomas (2006). "Comparing the EQ-5D and the SF-6D Descriptive Systems to Assess Their Ceiling Effects in the US General Population." Value in Health 9(4): 262-271.

Brown, M. M., G. C. Brown, et al. (1999). "Utility values and diabetic retinopathy." American Journal of Ophthalmology 128(3): 324-30.

Clarke, P., A. Gray, et al. (2002). "Estimating Utility Values for Health States of Type 2 Diabetic Patients Using the EQ-5D (UKPDS 62)." Med Decis Making 22(4): 340-349.

Clarke, P. M., A. J. Hayes, et al. (2009). "Using the EQ-5D Index Score as a Predictor of Outcomes in Patients With Type 2 Diabetes." Medical Care 47(1): 61-68 10.1097/MLR.0b013e3181844855.

Clarke, P. M., J. Simon, et al. (2006). "Assessing the Impact of Visual Acuity on Quality of Life in Individuals With Type 2 Diabetes Using the Short Form-36." Diabetes Care 29(7): 1506-1511.

Diener, E., E. M. Suh, et al. (1999). "Subjective well-being: Three decades of progress." Psychological Bulletin 125(2): 276-302.

Dolan, P. (1997). "Modeling Valuations for EuroQol Health States." Medical Care 35(11): 1095-1108.

Dolan, P., C. Gudex, et al. (1996). "The time trade-off method: Results from a general population study." Health Economics 5(2): 141-154.

Dolan, P., C. Gudex, et al. (1996). "Valuing health states: A comparison of methods." Journal of Health Economics 15(2): 209-231.

Dolders, M. G. T., M. P. A. Zeegers, et al. (2006). "A meta-analysis demonstrates no significant differences between patient and population preferences." Journal of Clinical Epidemiology 59(7): 653-664.

Farmer, A., A. Wade, et al. (2007). "Impact of self monitoring of blood glucose in the management of patients with non-insulin treated diabetes: open parallel group randomised trial." BMJ 335(7611): 132.

Glasziou, P., J. Alexander, et al. (2007). "Which health-related quality of life score? A comparison of alternative utility measures in patients with Type 2 diabetes in the ADVANCE trial." Health and Quality of Life Outcomes 5: 21.

Glasziou, P., J. Alexander, et al. (2007). "Which health-related quality of life score? A comparison of alternative utility measures in patients with Type 2 diabetes in the ADVANCE trial." Health and Quality of Life Outcomes 5(1): 21.

Grootendorst, P. V. (1997). "Health care policy evaluation using longitudinal insurance claims data: An application of the Panel Tobit estimator." Health Economics 6(4): 365-382.

Health Survey (1998), “Health Survey for England 1996”, The Stationery Office, London.

Heckman, J. J. (1979). "Sample Selection Bias as a Specification Error." Econometrica 47(1): 153-161.

Honore, B. E. (1992). "Trimmed Lad and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects." Econometrica 60(3): 533-565.

Huang, E. S., S. E. Brown, et al. (2007). "Patient perceptions of quality of life with diabetes-related complications and treatments." Diabetes Care 30(10): 2478-83.
Kind, P., G. Hardman, et al. (1999). "UK population norms for EQ-5D." Working Papers Centre for Health Economics, University of York November.
Levy, A., T. Christensen, et al. (2008). "Utility values for symptomatic non-severe hypoglycaemia elicited from persons with and without diabetes in Canada and the United Kingdom." Health and Quality of Life Outcomes 6(1): 73.
Lundkvist, J., C. Berne, et al. (2005). "The economic and quality of life impact of hypoglycemia." Eur J Health Econ 6: 197 - 202.
Lung, T., A. Hayes, et al. (2011). "A meta-analysis of health state valuations for people with diabetes: explaining the variation across methods and implications for economic evaluation." Quality of Life Research: 1-10.
Lykken, D. and A. Tellegen (1996). "Happiness is a stochastic phenomenon." Psychological Science 7(3): 186-189.
Matza, L., K. Boye, et al. (2007). "Utilities and disutilities for type 2 diabetes treatment-related attributes." Quality of Life Research 16(7): 1251-1265.
Mayou, R., B. Bryant, et al. (1990). "Quality of life in non-insulin-dependent diabetes and a comparison with insulin-dependent diabetes." Journal of Psychosomatic Research 34(1): 1-11.
Morgan, C. L., P. McEwan, et al. (2006). "Characterization and comparison of health-related utility in people with diabetes with various single and multiple vascular complications." Diabetic Medicine 23(10): 1100-1105.
Mullahy, J. (1998). "Much ado about two: reconsidering retransformation and the two-part model in health econometrics." Journal of Health Economics 17(3): 247-281.
NICE (2004). "Guide to the Methods of Technology Appraisal." (reference N0515).
Norman, G. R., J. A. Sloan, et al. (2003). "Interpretation of changes in health-related quality of life - The remarkable universality of half a standard deviation." Medical Care 41(5): 582-592.
Norman, G. R., J. A. Sloan, et al. (2003). "Interpretation of Changes in Health-related Quality of Life: The Remarkable Universality of Half a Standard Deviation." Medical Care 41(5): 582-592.
O'Reilly, D., F. Xie, et al. (2010). "Estimation of the impact of diabetes-related complications on health utilities for patients with type 2 diabetes in Ontario, Canada." Quality of Life Research: 1-5.
Ose, D., M. Wensing, et al. (2009). "Impact of Primary Care Based Disease Management on the Health-Related Quality of Life in Patients With Type 2 Diabetes and Comorbidity." Diabetes Care 32(9): 1594-1596.
Papke, L. E. and J. M. Wooldridge (1996). "Econometric Methods for Fractional Response Variables With an Application to 401 (K) Plan Participation Rates." Journal of Applied Econometrics 11(6): 619-632.
Papke, L. E. and J. M. Wooldridge (2008). "Panel data methods for fractional response variables with an application to test pass rates." Journal of Econometrics 145(1-2): 121-133.

Peeters, Y. and A. M. Stiggelbout (2010). "Health State Valuations of Patients and the General Public Analytically Compared: A Meta-Analytical Comparison of Patient and Population Health State Utilities." Value in Health 13(2): 306-309.
Powell, J. L. (1984). "Least absolute deviations estimation for the censored regression model." Journal of Econometrics 25(3): 303 - 325.
Pullenayegum, E. M., J.-E. Tarride, et al. (2010). "Analysis of Health Utility Data When Some Subjects Attain the Upper Bound of 1: Are Tobit and CLAD Models Appropriate?" Value in Health.
Redekop, W. K., M. A. Koopmanschap, et al. (2002). "Health-Related Quality of Life and Treatment Satisfaction in Dutch Patients With Type 2 Diabetes." Diabetes Care 25(3): 458-463.
Sakamaki, H., S. Ikeda, et al. (2006). "Measurement of HRQL Using EQ-5D in Patients with Type 2 Diabetes Mellitus in Japan." Value in Health 9(1): 47-53.
Sakthong, P., R. Charoenvisuthiwongs, et al. (2008). "A comparison of EQ-5D index scores using the UK, US, and Japan preference weights in a Thai sample with type 2 diabetes." Health and Quality of Life Outcomes 6(1): 71.
Sharma, S., A. Oliver-Fernandez, et al. (2003). "Utilities associated with diabetic retinopathy: results from a Canadian sample." British Journal of Ophthalmology 87(3): 259-61.
Sigelman, L. and L. Zeng (1999). "Analyzing Censored and Sample-Selected Data with Tobit and Heckit Models." Political Analysis 8(2): 167-182.
Sullivan, P., V. Ghushchyan, et al. (2008). "The impact of obesity on diabetes, hyperlipidemia and hypertension in the United States." Quality of Life Research 17(8): 1063-1071.
Sullivan, P. W. and V. Ghushchyan (2006). "Preference-Based EQ-5D Index Scores for Chronic Conditions in the United States." Medical Decision Making 26(4): 410-420.
Tabaei, B. P., J. Shill-Novak, et al. (2004). "Glycemia and the Quality of Well-Being in Patients with Diabetes." Quality of Life Research 13(6): 1153-1161.
UKPDS(37) (1999). "Quality of life in type 2 diabetic patients is affected by complications but not by intensive policies to improve blood glucose or blood pressure control (UKPDS 37). U.K. Prospective Diabetes Study Group." Diabetes Care 22(7): 1125-1136.
Walters, S. J. and J. E. Brazier (2005). "Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D." Quality of Life Research 14(6): 1523-1532.
Weinberger, M., M. Kirkman, et al. (1995). "A nurse-coordinated intervention for primary care patients with non-insulin-dependent diabetes mellitus." Journal of General Internal Medicine 10(2): 59-66.
Welch, G. W., A. M. Jacobson, et al. (1997). "The Problem Areas in Diabetes Scale. An evaluation of its clinical utility." Diabetes Care 20(5): 760-766.
Wit, G. A. D., J. J. V. Busschbach, et al. (2000). "Sensitivity and perspective in the valuation of health status: whose values count?" Health Economics 9(2): 109-126.

Wooldridge, J. M. (1995). "Selection corrections for panel data models under conditional mean independence assumptions." Journal of Econometrics 68(1): 115-132.
Wooldridge, J. M. (2002). Econometric Analysis of Cross Section and Panel Data, MIT Press.

Chapter 3

THE IMPACT OF DIABETES-RELATED COMPLICATIONS ON HEALTHCARE COSTS

Abstract

Reliable estimates of the impact of complications on resource use and healthcare costs are important for researchers studying cost-effectiveness of interventions. This chapter shows new estimates of the cost of diabetes complications using 10 years of additional data from the landmark UK Prospective Diabetes Study (UKPDS). I analyse inpatient and non-inpatient utilisation, using hospitalization records for England for 2,791 patients and six non-inpatient resource use questionnaires administered to UKPDS participants between 1997 and 2007 resulting in 3,589 responders and 10,920 questionnaires. Resource use measured includes home, clinic, and telephone contacts with general practitioners, nurses, podiatrists, opticians, and dieticians, eye and other hospital out-patient clinics, and all in-patient stays and procedures.

Herein I estimated the immediate and long term impact on impatient and non-inpatient costs of myocardial infarction, ischemic heart disease, stroke, heart failure, amputation, blindness in one eye, retinal photocoagulation, cataract extraction and vitreous haemorrhage, controlling for patient specific characteristics.

Non-inpatient costs account for roughly 30% of the total expenditure identified, but routine outpatient care costs increased substantially for patients in the study over the 10-year follow-up, and particularly between 1997 and 2003. This period corresponded with the publication of the main UKPDS trial results demonstrating the benefits of intensive therapy and the subsequent changes in treatment guidance and protocols.

Amputations, ischemic heart disease, myocardial infarction and stroke are found to be the highest drivers of cost both within the year in which they occur but also in subsequent periods.

3.1 Introduction

Reliable estimates of the impact of complications on resource use and healthcare costs are important both in order to understand the burden of a condition and to estimate the cost-effectiveness of interventions that might reduce complications. Both objectives coincide as the cost estimate of the burden of a disease is intended to give information on the value that can be gained from reducing the incidence and prevalence of a specific complication. When evaluating therapies for chronic conditions, this usually involves estimating the potential savings that accrue from reducing complications associated with the disease (Gold MR 1996).

The costs associated with a disease are normally divided into direct and indirect costs. Direct costs are those arising within the health care system, i.e., medical costs for diagnoses, treatment, drugs, rehabilitation and prevention. In the UK the direct cost of diabetes is estimated to represent 10% of the NHS budget (£9 billion a year)[1] and this cost is expected to increase by up to 30% in the next three decades (Bagust, Hopkinson et al. 2002). Considering that there are currently an estimated 3 million people living with diabetes in the UK, this translates into £3,000 per head in direct costs for the overall population with diabetes. The available evidence indicates that patients with diabetes consume two to six times more resources than individuals of similar age, gender and ethnicity (Norlund, Apelqvist et al. 2001) who do not have diabetes.

Indirect costs represent the value of lost output caused by production losses, i.e., costs generated by sick leave, early retirement and premature death. Other

[1] http://www.diabetes.org.uk/Documents/Reports/Diabetes_in_the_UK_2010.pdf

components of indirect costs are considered to be the burden a condition may impose on third parties (in particular family and friends), and other social consequences. However the methods for converting these losses into monetary costs are more complicated.

Despite the importance of diabetes for healthcare systems, very few economic studies have considered the individual consumption of healthcare resources. Most of the studies available make estimates of healthcare consumption and costs from data on population prevalence and as such do not report the clinical factors that increase the cost of the condition. When individual patient data are available, healthcare costs are often calculated using administrative prices or charges. It is important however to go beyond pure accounting costs for two main reasons. First, we want to be able to estimate the increase in all healthcare costs in the year in which the complication occurs, but also on costs in the years after a patient has a complication. Second, average costs may mask a lot of patient heterogeneity: the underlying level of healthcare utilization is often a phenomenon endogenous to the patient and his/her characteristics and for given applications we might want to disentangle this from the cost of the complication itself.

The focus of this chapter is on a subset of direct costs, those that accrue as a result of inpatient and day care as well as outpatient and other non-inpatient resource utilization. This captures most treatment and management costs, but reliable information was not available on the costs of medications prescribed in general practice and so these were excluded. The point of view adopted here is that of a single payer system, the NHS. In-patient and day care costs are estimated using multiple regression analysis based on costs calculated from Healthcare Resource Groups (HRG) on all English

patients who took part in the Post Trial Monitoring phase of the study from October 1998 onwards. Non-inpatient medical expenditures were estimated using questionnaires completed by individuals who participated in the UKPDS follow up study (1997- 2007) and provided information on all home, clinic, and telephone contacts with doctors, nurses, podiatrists, opticians, and dieticians, and with eye and other hospital out-patient clinics over recall periods varying from 4 months to one year. Although the bulk of total medical expenditures are accounted for by hospitalization costs, the management of patients with diabetes has shifted over time to greater outpatient utilization. From these sources, estimates are made of the effects of six pre-specified clinical events on health care costs: fatal and non-fatal myocardial infarction (MI), fatal and non-fatal ischemic heart disease (IHD), fatal and non-fatal stroke, heart failure, amputation, and blindness in one eye.

This chapter builds on previous published work using data on diabetes-related complications from the UKPDS (Clarke, Gray et al. 2003). In that earlier study, inpatient episodes occurring up to 1996 were costed by obtaining information on the specialty in which the admission occurred, and attaching a specialty-specific cost per inpatient day using information from Trust Financial Returns (TFR2)[2]. That approach was relatively simple to implement but also had limitations. For instance, attaching a flat rate per day to the number of bed days may not reflect accurately the cost of a hospitalization, as it is likely that the bulk of costs fall at the beginning of the hospitalization period. Similarly,

[2] Information on hospital in-patient stays was collated by the UKPDS based on patient interviews at every clinic visit while the patient participated in the trial, death records and hospital records. Where an in-patient stay had occurred, details were obtained from the relevant hospital of dates of admission and discharge, reasons for admission, and any major procedures undertaken. All hospital episodes were subsequently allocated to one of 40 national specialty codes using information from the hospital or assessment by a panel of clinicians. In approximately 16%of cases (predominantly in the earliest years of the study) the length of stay was not recorded and so multiple imputation methods were applied to replace the missing data.

the cost of any given procedure for a patient without co-morbidities is likely to be considerably lower than the cost of the same procedure if the patient has co-morbidities (Zsolt and Smith 2005). To address these issues, a HRG-based method to cost admissions (HRG4) that takes into account not only specialty and length of stay but also other characteristics of the patient and the admission is used. This method is not only more accurate but it also reflects the current payment mechanism in England (Geue, Lewsey et al. 2011). In terms of non-inpatient costs, UKPDS 65 analyzed the first questionnaire which had a recall period of 4 months. The data were then annualized as the research interest is in annual costs. The annualization was done by multiplying resource use by 3 and thus assuming that the cost flow is uniformly distributed within a year. The authors then defined a short and long run scenario by looking at events occurring one year before the questionnaire date and all periods previous to the one year but did not have enough power to differentiate costs associated with acute and long term events.

This chapter is divided into seven sections. Section 3.2 outlines the general approaches used in the literature to model healthcare costs and summarizes the cost of diabetes-related complications on inpatient and non-inpatient resource use as reported in published studies. Section 3.3 illustrates the data used, in particular HES (Hospital Episode Statistics) and HRGs (Healthcare Resource Groups). This section also describes the questionnaires on outpatient and other resource utilization administered to UKPDS participants between 1997 and 2007.

Section 3.4 addresses a number of methodological issues pertinent to both inpatient and non-inpatient cost data, namely: (1) functional form issues (skewness and probability mass accumulating at zero); (2) differences between questions of incidence and questions of causality and the situations in which both may coincide; (3) the

possible presence of fixed effects in healthcare costs (i.e., is there an intrinsic level of resource utilization which depends on type-- patient specific characteristics and preferences – and exists independently from the occurrence of events?);

Section 3.5 outlines the modelling strategy chosen. Section 3.6 presents the results, model checks and tests. Finally, section 3.7 discusses the contributions and shortcomings of the findings as well as the implications of the results.

3.2 Literature

Many costing studies have documented the total burden of diabetes to the health system and society, with varying focus on the different components of costs (Pagano et al. 1999). However, only a handful of studies have reported cost estimates for a range of particular diabetes-related complications.

Estimates of the cost of diabetes to the NHS were first attempted in 1989 by Gerard et al. who found that 4-5% of all UK healthcare expenditure was spent on people who had coexisting diabetes (Dixon, Currie et al. 2000). 20 years later, as much as 12.6% of healthcare expenditure has been claimed to be attributable to diabetes (Morgan, Peters et al. 2010).

Population models (Bagust, Hopkinson et al. 2002; ADA 2003) that calculate the overall burden of diabetes rely on prevalence rates to estimate the total economic burden of the disease to which they attach costs. There are two potential shortcomings with this: 1) values might reflect the diagnosed population rather than the actual number of people with diabetes; 2) underreporting of diabetes in health care databases is likely to occur when other co-morbidities are present. In the case of T2DM, patients tend to have multiple pathology and multiple risk associations. Because at the

population level there is no single data source that contains all the information necessary to estimate the health care cost, it is therefore necessary to draw upon multiple data sources. Among some of the issues related to using different data sources are differences in definitions for identifying people with diabetes and differences in levels of detail available. It is thus understood that the use of aggregate figures on resource consumption, also known as the "top down approach" is likely to be a very conservative approach for calculating costs (Dixon, Currie et al. 2000). This is why the "bottom-up" approach is usually favoured, starting from a selected subpopulation with the actual disease, results can then be extrapolated to the national level. Often however a combination of the two approaches is used when estimating the cost-of-illness of a disease. Furthermore, if the goal is to curb the cost of diabetes, it is important to understand what the main drivers of costs are. Diabetes complications have been shown to be an important contributor to the high healthcare costs associated with managing the condition (Jönsson, 2002).

This section is divided into two parts: the first summarizes the evidence available in the literature and seeks to provide references for ranges as well as rank the impact of different types of complications on hospital costs (inpatient and day cases). The second part focuses on non-inpatient resource utilization.

Using PubMed and the Wiley online library as the main bibliographic data bases, the following terms were searched for: healthcare resource utilization, T2DM, inpatient, non-inpatient, outpatient costs, diabetes-related events and complications. The final search was completed in October 2012. The aim of the review was both methodological and restricted to studies looking at the impact of diabetes-related complications on costs only rather than costs associated with the condition as a whole. This restriction

was relaxed for non-inpatient studies to accommodate for the fact many studies failed to find statistically significant decrements on costs for individual complications and to allow for a wider review on statistical methods as the evidence on non-hospital costs is scant. Because the nature of outpatient resource utilization has changed dramatically over the past decade, the review only considers studies published post 1998.

Inpatient costing studies in Diabetes

Since 1998, six studies have reported the impact of diabetes complications on hospitalization costs using patient level data (O'Brien, Shomphe et al. 1998; Clarke, Gray et al. 2003; Clarke, Leal et al. 2008; Gerdtham, Clarke et al. 2009; Clarke, Glasziou et al. 2010; Fu, Qiu et al. 2010). These are mostly European studies; the exception being Clarke et al. (2008) who looked at the cost of complications in Australia using administrative health care data and O'Brien (2008) who estimated the cost of diabetes-complications in the US. All studies but two used Generalized Estimating Equation (GEE) approach to estimate hospital costs for diabetic patients and assumed a gamma function in order to describe the relationship between variance and conditional mean. They all assumed a log link function, meaning they all assumed that covariates act multiplicatively on the mean. Two studies (Clarke, Leal et al. 2008; Gerdtham, Clarke et al. 2009) used fixed-effects (FE) estimators, alongside a GEE model to aid interpretation as FE estimates additive impacts of complications on costs.

Previous studies of diabetes-related complications have shown that costs have a distinct profile over time. For instance Brown et al. (1999) analyzed 9 years of clinical

data on 11,768 members of a large health maintenance organization who had probable[3] type 2 diabetes mellitus. They looked for the presence of cardiovascular and renal complications, staged their progression, estimated their incremental costs by stage and found very large increments in resource utilization over time.

The majority of studies in this review estimated both the immediate or acute impact of the first diabetes-related complication (i.e., the effect in the year the complication occurs) and the long-term impact (i.e., the effect in all subsequent years), which is also often described as state costs or history costs. A few studies find evidence of the impact of history of events on current expenditure (O'Brien, Shomphe et al. 1998; Clarke, Gray et al. 2003) but most are underpowered to detect differences between the short and long term effects. Throughout all the estimates available, amputations and cardiovascular disease are associated with the highest costs (Table 3.1).

[3] The authors used the Kaiser Permanente Northwest Division (KPNW) registry. Members were presumed to have type 2 diabetes mellitus if they entered the registry after age 45 years or entered earlier and had insulin dispensed within 2 years of entry.

Table 3.1: Review of studies looking at the impact of diabetes-related complications on hospital costs

Author (year)	Country	Subjects	Cost Source	Methods	Effects
Clarke (2003)	UK	3,488 patients in the UKPDS	Hospital length of stay (LOS) multiplied by TRF specialty per diem costs	2PM. Multiple regression analysis (random effect logit in the first part, GLM gamma with log link in the second part)	Amputation £8,459 (95% CI: £5,295, £13,200); nonfatal MI £4,070 (£3,580, £4,722); fatal MI £1,152 (£941, £1,396); fatal stroke £3,383 (£1,935, £5,431); non-fatal stroke £2,367 (£1,599, £3,274); ischemic heart disease £1,959 (£1,467, £2,541); heart failure £2,221 (£1,690, £2,896); cataract extraction £1,553 (£1,320, £1,855); and blindness in one eye £872 (£526, £1,299). The annual average in-patient cost of events in subsequent years ranged from £631 (£403, £896) for heart failure to £105 (£80, £142) for cataract extraction
Clarke (2008)	Australia	70,340 patients with diabetes	administrative casemix health care data	Multiple regression analysis (GLM gamma with log link) and FE	For a man aged 60 years, the average costs in the year the event first occurred were: amputation $20,416 (95% CI 18,670–22,411); nonfatal MI $11,660 (10,931–12,450); nonfatal stroke $14,012 (12,849–15,183); IHD $12,577 (12,026–13,123); heart failure $15,530 (13,965–17,009); renal failure $28,661 (22,989–34,202); and chronic leg ulcer $15,413 (13,089–18,123).

Clarke (2010)	Asia, Eastern Europe, and Established Market Economies	11,140 patients in the ADVANCE study	LOS multiplied by regional estimates of the costs per day	Multiple regression analysis (GLM gamma with log link)	Probabilities of hospitalization were lowest in Asia and highest in Established Market Economies; but LOS were highest in Asia and lowest in Established Market Economies. Estimated annual hospital costs for patients with none of the specified events or event histories ranged from $76 in Asia to $296 in Established Market Economies. All complications included in this analysis led to significant increases in hospital costs; coronary events, cerebrovascular events, and heart failure were the most costly, at more than $1,800, $3,000, and $4,000 in Asia, Eastern Europe, and Established Market Economies, respectively.
Fu (2010)	Spain, France, UK, Norway, Finland, Germany and Poland.	Real-Life Effectiveness and Care Patterns of Diabetes Management study	Data from interviews with patients	Logit to identify the relationship between concurrent MVC and the likelihood of ER admission, surgical procedures, and hospitalization. Negative binomial models used to predict the number of office visits and length of hospital stay per year attributable to the concurrent MVC	Within a seven-country European sample, this study showed that T2DM patients with MVC were more likely to use healthcare resources compared with T2DM patients without MVC.

Gerdtham (2009)	Sweden	179,749 patients	administrative hospital casemix-based cost data (NordDRG)	Multiple regression analysis (linear (FE) and GEE (gamma family and log link)	The average annual costs (linear model) associated with inpatient admissions for a 60-year-old male in the year the first event first occurred were as follows: €6,488 (95% CI 5,034; 8,354) for diabetic coma; €6,850 (95% CI 6,514; 7,204) for heart failure; €7,853 (95% CI 7,559; 8,144) for non-fatal stroke; €8,121 (95% CI 7,104; 9,128) for peripheral circulatory complications; €8,736 (95% CI 8,474; 9,001) for non-fatal myocardial infarction (MI); €10,360 (95% CI 10,085; 10,643) for IHD; €11,411 (95% CI 10,298; 12,654) for renal failure; and €14,949 (95% CI 13,849; 16,551) for amputation
O'Brien (1998)	USA	709,108 discharges of patients with type 1 or type 2 diabetes (out of 8 million)	unit costs from acute care discharge databases, clinical guidelines, fee schedules, and the literature	Accounting costs	Acute myocardial infarction ($27,630 event cost; $2,185 state cost), generate a greater financial burden than do early-stage complications, such as microalbuminuria ($62 event cost; $14 state cost). Yet, complications that are initially relatively low in cost (e.g., microalbuminuria) can progress to more costly advanced stages (e.g., end-stage renal disease, $53,659 state cost)

Non-inpatient costing studies in Diabetes

The data collection methods used for outpatient and other costs that do not require overnight hospital stays, vary from claims databases and administrative billing databases– which are the predominant source to quantify utilization – to patient-reported questionnaires (Clarke, Gray et al. 2003; daCosta, Cappelleri et al.) and physician-reported questionnaires (Gandjour, Kleinschmit et al. 2002; Schmitt-Koopmann, Schwenkglenks et al. 2004).

The most common method used is analysis of variance to test differences in costs by age and racial groups (Maciejewski and Maynard 2004); and those with and without specific conditions such as severe hypoglycaemia (Nichols, Arondekar et al. 2008; Hammer, Lammert et al. 2009), cardiovascular disease and chronic kidney disease (Levin, Chaudhry et al. 2009). Only five studies out of the thirteen studies indentified by the search, explicitly address functional form issues. Bogner et al.(2010) model costs on the logarithmic scale, Clarke (2003), DaCosta(2011), and Gilmer (2005) use a gamma distribution to specify relationship between conditional mean and the regressors.

Beyond methods employed, it is difficult to compare studies and their findings as the "basket of good" being analyzed differ across the studies. Some include outpatient costs only (GP, PCP[4] and nurses), others are more broadly non-inpatient (including eye clinics and other outpatient facilities), and some include the cost of drugs in the estimation of non-hospital costs (Ballesta, Carral et al. 2006). Only four studies looked at the impact of specific diabetes-related complications on costs. It is also important to point out while comparing results across countries, that context itself might be a driver of differences in utilization. Practices in the management of disease may differ across

[4] Primary Care Providers

health care systems. Payment systems that use fee for service (FFS) (rather than capitation or salary) are likely to have higher costs because they favour specialization over generic primary care functions. In such contexts claims are made for every service provided, which also might raise administrative costs (Berenson and Rich 2010). Lastly, physicians who are paid through FFS have the incentive to create supplier induced demand. Levin et al. (2009) point out that while patients with chronic kidney disease (CKD) comprise only 7.5% of diabetes patients in British Columbia, 34% of these generate duplicate laboratory tests. In 2005, there were over 747,000 duplicate laboratory tests done for this subset of patients, costing almost three million dollars in unnecessary costs.

All studies (unless they analyze a single specific co-morbidity as DaCosta(2011)) were underpowered to obtain estimates for several specific complication. The most common approach was to group complications into macro-vascular (MI, stroke, IHD, heart failure) and micro-vascular (eye related and renal problems). The overall outpatient expenditure represented one fifth to one third of total direct medical expenditure.

Table 3.2: Review of studies looking at the impact of diabetes-related complications on outpatient resource utilization

Author(s), publication year	Country or trial population	Data sources	Statistical Method	Effect/ Findings
Ballestra et al. (2006)	Spain. 1 year of patient level data	A descriptive observational study conducted in a representative sample of 517 patients with type II diabetes. Direct costs include medication, monitoring tests, medical and nursing consultation, cost of treatment of complication, emergency attendance and hospitalizations. Indirect costs were calculated for productivity loss attributable to diabetes during the study year on the basis of work days lost.	Three forward stepwise multivariate regression models were constructed, taking the direct, indirect, and total costs as the dependent variable. No indication of the functional form used is given.	A total annual health cost of €4,278/patient was calculated (direct €2,504; indirect €1,774)
Bogner (2010)	USA. Health care expenditure for elderly Medicare enrolees with DM.	A retrospective case-control identifies 4 groups of elderly patients with HF and DM (5,498), HF only (51,089),DM only (5,971), and no-HF and no-DM (55,438) using an administrative database. Health resource utilization, and cost (reimbursement) data were obtained from the Medicare claims database for the years 2000 and 2001	Analysis of variance to compare means and regression analysis using logs and HR.	The mean total costs was highest for the group with HF and DM ($32,676), and second highest for the HF only group ($22,230). In multivariable model that adjusted for potentially influential covariates, the group with HF and DM had a 3-fold higher cost than the group without DM and HF 95% CI 3.82; 5.31).
Clarke (2003)	UK (UKPDS)	A cross-sectional survey of 3,488 UKPDS patients' outpatient healthcare usage (all home, clinic and telephone contacts with GPs, nurses, podiatricians and dieticians and with eye and other hospital out-patient clinics) conducted in 1996–1997.	Regression equations to derive annual probability of incurring non-in-patient costs, and costs of non-in-patient care conditional on incurring a cost (Part 1 equation derived from logistic regression, Part 2 equations derived from GLM with gamma distribution, and logarithmic link function)	Non-in-patient costs for macrovascular complications were £315 (£247, £394) and for microvascular complications were £273 (£215, £343) in the year of the event. In each subsequent year the costs were, respectively, £258 (£228, £297) and £204 (£181, £255)

DaCosta (2011)	USA. three consecutive waves of the National Health and Wellness Survey (2006–2008)	Participants in the national survey were categorized into (1) those with pDPN (painful diabetic peripheral neuropathy) (N = 290), (2) those with diabetes but without pDPN (N = 1,037), and (3) those not diagnosed with diabetes ("control group"; N = 8,162). Direct costs were calculated for each of the three groups by multiplying the number of ER visits, hospitalizations, and physician visits to the the MEPS database costs.	Generalized estimating equations (GEE). A normal distribution was specified for health status outcomes, and a negative binomial distribution was specified for the work productivity, physician visit, and ER visit outcomes. As a post hoc analysis, pDPN patients were matched to controls using 1:1 propensity score matching	The pDPN group visited significantly more health care providers than the group without pDPN (5.69 vs 4.76, P-value=0.026). The difference between the pDPN group and the control group was not significant (5.69 vs 4.93, P-value=0.10).
Fu (2010)	Europe. Multicenter observational study in Spain, France, UK, Norway, Finland, Germany and Poland.	A control cohort with T2DM but without concurrent macrovascular complications (MVC) was identified using 1:1 propensity score matching.	Logit models used to identify the relationship between concurrent MVC and the likelihood of ER admission, receiving medical/surgical procedures, and hospitalization. Negative binomial models were used to predict the number of office visits and length of hospital stay per year attributable to the concurrent MVC.	Relative to controls, patients with MVC were significantly more likely to have ER admissions [odds ratio (OR) 2.69; 95% CI: 1.56–4.65], receiving medical/surgical procedures (OR 2.57; 95% CI: 1.56–4.21) and hospitalizations (OR 2.58; 95% CI: 1.64–4.07) Similarly, MVC were associated with 1.49 additional office visits per year (p = 0.036) and 0.32 days of hospital stay per year (p = 0.023).

Gandjour (2001)	France, Germany, Italy, The Netherlands, Sweden, Switzerland, and the UK	188 European physician practices assessed annual services for one hypothetical average patient. In countries with a detailed fee-for-service schedule (Germany, Italy, and Switzerland) reimbursement fees were used to approximate costs and their averages were use to impute costs for other countries. For each country an average quality rating was calculated by weighting the response to each quality indicator with the level of scientific evidence.	Analysis of variance to test for differences between the country-specific average quality ratings and to test for differences in resource utilization.	Average quality ratings ranged from 0.40 in The Netherlands to 0.62 in the UK (0 = lowest rating; 1 = highest rating). Total annual costs for secondary prevention were higher in Switzerland than in Germany and Italy (€475, €381, and €283, respectively). Resource utilization was highest in Germany and lowest in the UK.
Garattini (2004)	Italy. Multicenter, prospective, observational study involving 31 Italian diabetes centres	A total of 1,910 patients was classified into eight prognostic groups by type of diabetes (types 1 and 2), metabolic control (HbA1c >7.5%,HbA1c ≤7.5%) and age (≤60, >60).	CIs for hospitalizations were based on the Poisson distribution. A nonparametric test (Kruskal-Wallis) was used to assess differences in costs for patients with type 1 and type 2, and patients with poor (HbA1c >7.5%) and good (HbA1c≤7.5%) metabolic control.	The average total cost of type 1 diabetes per patient per year ranged from € 762 in group 2 (age ≤60, HbA1c >7.5%) to € 1,060 in group 4 (age >60,HbA1c >7.5%). The cost of type 2 diabetes from € 423 in group 5 (age ≤60, HbA1c ≤7.5%) to € 613 in group 8 (age >60, HbA1c >7.5%).
Gilmer (2005)	USA. Survey and medical claims in an diabetic cohort	Data from a patient survey (1,694 adults with diabetes) and medical record reviews merged with 3 years of medical claims	Costs were modelled using multivariate GLM regression with gamma distribution, and log link function	3-year costs in those with coronary heart disease (CHD) and hypertension were over 300% of those with diabetes only ($46,879 vs. $14,233; P 0.05). Depression was associated with a 50% increase in costs ($31,967 vs. $21,609; P 0.05). Relative to those with a baseline A1c of 6%, those with an A1c of 10% had 3-year costs that were 11% higher ($26,408 vs. $23,873; P 0.05).

Nichols (2008)	USA. patients with impaired fasting glucose (IFG), impaired glucose tolerance (IGT), or both	26,111 non-diabetic patients >45 in HMO, assigning them to different glucose categories	Mean annual medical resource utilization and age/sex-adjusted costs over one year were compared	Total outpatient visit utilization was similar across all 4 groups but inpatient costs greatly differed with patients with higher hyperglycemia being 20.3% more likely to be hospitalized.
Hammer (2009)	Germany, Spain and the UK. Diabetes patients with severe hypoglycaemic events (SHEs)	Healthcare resource use was measured by surveying 639 patients aged ≥16 years, receiving insulin for type 1 (n=319) or type 2 diabetes (n=320), who experienced ≥1 SHEs in the preceding year. Patients were grouped by location of SHE treatment: group 1, community (family/domestic); group 2, community (healthcare professional); group 3, hospital. Costs were calculated from published unit costs applied to estimated resource use. Costs of resource use included in the analysis: primary care physicians (PCP) office visits (€7.40 in Germany, €15.10 in Spain, £32.42 in UK); PCP phone call (€7.40 in Germany, €10.11 in Spain, £28.23 in UK); diabetes specialist (€19.40 in Germany, €52.35 in Spain, £112.93 in UK)	Average costs	Average costs were lower for type 1 than for insulin-treated type 2 diabetes, in all three countries. Telephone calls, visits to doctors, blood glucose monitoring and patient education contributed substantially to costs for non-hospitalised patients, including those in hospital (group 3) in all 3 countries.
Schmitt-Koopman et al (2004)	Switzerland. Survey	111 primary care physicians from 19 cantons throughout Switzerland and their 1479 patients with diabetes.	Non-stratified and stratified descriptive analyses	Patients with diabetes on average had 10.3 consultations per year related to this disease (95% CI: 10.0–10.7). Direct costs were presented in aggregated form separately for outpatient, inpatient care and medications.

Levin et al (2009)	Canada	Administrative billing databases (outpatient visits, laboratory tests, pharmacy and hospital inpatient services) over 3 years (2003–2005)	itemized a triad of conditions (cardiovascular disease, DM and chronic kidney disease) to provide averages for resource utilization associated with each condition and combinations	CKD, CVD and DM diagnoses are found in 422 124 persons within a province of 4.3 million individuals (10%); 1.7% had all three conditions. The median number of physician visits was 26 per patient year. Duplicate testing accounted for expenditures of $3 million/annum; 7.55% of patients accounted for 34.4% of duplicate tests. Those with DM or CKD had similar use of medications, physician visits and hospital days.
Maciesjewski (2004)	USA. Veterans with diabetes receiving care in Veterans Administration (VA) facilities.	Number of inpatient hospitalizations and outpatient clinic visits in 1998 by veterans with diabetes receiving care in Veterans Administration (VA) facilities.	mean in utilization across different age and racial groups, as well as total direct inpatient and outpatient costs	The average number of outpatient visits was 4.5 in 1997 and 4.6 in 1998. VA incurred $214.8 million in outpatient expenditures in 1998 for a population of 429,918 veterans

3.3 The data: HES, HRG, Costs and Complications

From September 1997 the intervention phase of the UKPDS study closed and 4,189 surviving patients entered a post-trial monitoring study, which continued to collect information on clinical events and resource use until September 2007, while participants were treated to best practice by their GPs. Details of the study design and the main clinical results have been published elsewhere (UKPDS(33) 1998; UKPDS(34) 1998; UKPDS(38) 1998).

Of the initial 5,102 patients recruited to the UKPDS trial, 1,724 (24.75%) did not live in England (1,263 in Scotland and 461 in Northern Ireland). Scottish and Northern Irish survivors were therefore not included in the analysis of inpatient costs as they are not expected to appear in English hospital admission records (which are the source of inpatient resource use and costs in this sub study). Our estimation sample starts on the 1st October 1998 and concludes on the 30th of September 2007 corresponding to the ending of the UKPDS post-trial monitoring period. Conditional on the patient being alive on 1st October 1998, we expected to match 3,074 English patients to the English HES data set.

This section is divided into 4 parts. Hospital Episode Statistics (HES) are described first, followed by Healthcare Resource Groups (HRG), derived costs and finally the incidence of complications and other characteristics of patients studied.

Hospital Episode Statistics (HES)

England is unusual in having made it mandatory for all NHS hospitals to report activity hospital episode statistics[5]. In very few countries are disaggregated cost data routinely available (Schreyögg, Stargardt et al. 2006). This means that analysis of hospital costs in other countries is often based on either a limited sample of hospitals or utilises data reported at hospital level, such as total expenditure.

In this thesis a record linkage between the UK Prospective Diabetes Study (UKPDS) and the Hospital Episode Statistics (HES) databases was conducted. This was possible by linking participants' identifiers from the UKPDS study to the HES dataset held by the Unit of Health-Care Epidemiology (UHCE) at the University of Oxford. The UHCE had already built a linked HES dataset for England spanning the period 1998 - 2011. The UKPDS patients were identified by using information on their NHS numbers, date of birth, and postcode which was provided by the Diabetes Trial Unit (DTU) to the UHCE and managed through the DTU Standard Operating Procedures governing the export and security of data. The UHCE then extracted all hospitalisation records related to the UKPDS patients during the post-trial monitoring study. The linkage was done using encrypted NHS numbers (cross checking matches with information on gender, ethnicity, dates of birth and postcodes) and requesting the Health and Social Care Information Centre UK to use the same encryption methods on the UKPDS identifiers as it used in the UHCE HES dataset. The HES records that matched the encrypted UKPDS identifiers were sent to the DTU's UKPDS database.

After the removal of all patient identifiable variables, the dataset was provided to me for data analysis in order to estimate the impact of diabetes' complications on

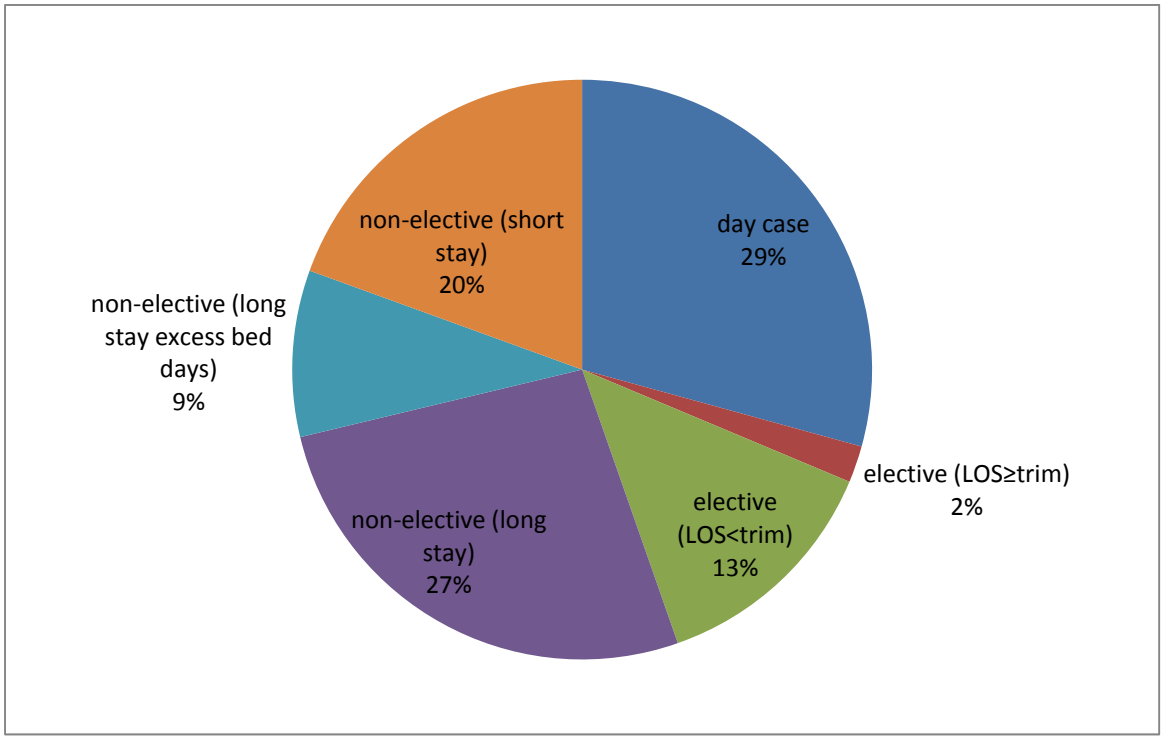
[5] Hospital Episode statistics do not measure costs directly.

costs. The linking of two already established and approved databases (UKPDS and HES) enabled me to obtain the required data for the analysis, while preserving the anonymity of the patients, and minimising the risk of security breaches.

It is important to highlight the special ethical considerations necessary to make this research possible. Formal consent was obtained from participants at enrolment by way of face-to-face interviews and from their GPs via questionnaires. Given that a significant number of UKPDS participants had died during the follow up period, it would have been impossible to re-confirm the patients' consent to use their linked HES data. The omission of those who have died or might have become untraceable would have substantially biased the study. Approval was thus obtained from the ethics committees' at all 23 clinical centres participating in the UKPDS, and it was sanctioned that this study conforms to the Declarations of Helsinki guidelines.

2,791 of the 3,074 UKPDS patients entering the post-trial monitoring study in England were found in the HES records. These patients contributed to 27,489 hospital spells over 9 years. More than half of all hospital spells were non-elective (55%). 29% were day cases and the remaining 15% were elective hospitalizations.

Figure 3.1: Hospitalization type



LOS=length of stay. **Trim**= trim point – bed days that fall outside nationally set lengths of stay for any given HRG.
Daycase: A patient admitted electively (i.e. from a waiting list) during the course of a day with the intention of receiving care without requiring the use of a hospital bed overnight.
Elective admission: A patient admitted from a waiting list.
Non-Elective admission: A patient not admitted from a waiting list i.e. either admitted as an Emergency (e.g. A&E), Maternity or "Other" (e.g. transfers other than in an emergency). The reference cost guidance requires all non-elective patients should be separately identified as either **short stay** (the average length of stay is less than two days) or **long stay** (greater or equal to two days). **Excess bed days** are days above HRG-specific trim points.

The median hospital episode in the sample had a duration of 2 days across any year. The average length of stay in hospital per spell was 13 days. A day in hospital is counted when an over-night stay has occurred (day cases are coded as zeros). This information is summarized in Table 3.3 for each one of the years under consideration. Year one starts on the 1st of October 1998. While the length of stay in hospital did not change substantially over time, the number of hospitalizations more than quadrupled in the 9 years analyzed.

Table 3.3: Duration and Number of Hospitalizations over time

Year	1	2	3	4	5	6	7	8	9	All
Length of Stay in Hospital										
Mean	11.2	11.9	12.4	13	14.5	14.6	13.6	15.3	14.2	13.3
SD	23	24.3	25.2	24.6	28.1	28.3	27.9	28.7	28	26.4
Max*	212	196	195	205	209	204	193	207	204	212
Number of Hospitalization Spells										
Mean	2.1	2.2	2.3	2.5	2.5	2.8	2.9	2.9	2.8	2.5
SD	2.2	2.8	2.9	2.7	2.5	6.1	5.6	5.3	3.8	3.9
Median	1	1	2	2	2	2	2	2	2	2
Max	32	53	59	42	22	144	96	120	52	144

*min is always zero for LOS and median is 2 throughout all years.

Healthcare Resource Groups (HRGs)

In England, the Reference Costs form the basis for calculating the national tariff, according to which hospitals are paid under Payment by Results (PbR). Hospitals in England do not collect detailed information on each patient's resource use in a way that would allow for bottom-up costing of each individual's patient care. Rather, every hospital applies a standard top-down costing methodology to produce a Reference Cost for each elective, day case and non-elective HRG in each specialty. This means that total hospital costs are progressively cascaded down first to treatment services (wards, theatres, etc), then to specialties, and finally to HRGs. Reference Costs are calculated on a full absorption basis, meaning that they are meant to reflect the full and true cost of the service delivered (Department of Health, 2008).

Before the onset of HRGs, Foundation Trusts (FTs) were required to provide information on total expenditure. These so called Trust Financial Returns (TFR) were grouped into specialties, which were publicly available up to the April 2003- 2004 financial year when they were supplanted by the Healthcare Resource Group. The

existing published work on UKPDS costs of complications used TFR (Clarke, Gray et al. 2003).

Since 2004 a Payment by Results (PbR) system of reimbursement for providers was set up. Under PbR, contracts are struck between commissioning bodies and providers on the basis of predicted levels of activity adjusted for casemix (BMA March 2008). Casemix refers to the way in which different diagnoses and treatments together with their different resource needs are taken into account. This system classifies types of patients treated, and costs each category to enable consistent pricing (Palmer 1996). Under the casemix approach to payment, hospitals are funded on the basis of the numbers and types of patients they treat. Currently, every hospital in England applies a standard top-down costing method to produce a Reference Cost for each specialty. These returns incorporate overheads and are widely accepted as reliable indicators of hospital service costs (Laudicella, Rosen Olsen et al. 2009).

A similar method of categorisation has been used in the United States since 1983, to reimburse hospitals for the services provided to inpatients under the Federal government's Medicare programme. Hospitalization episodes in the US are classified into Diagnosis Related Groups (DRGs)[6]. This type of financing of hospital inpatient activities is known as 'output-based' funding. Robert Fetter (the creator, along with John Thompson, of the DRG system) has suggested that this conceptualisation of diagnostic

[6] Interestingly, major teaching hospitals in the United States were extremely active in opposing the American variant of casemix funding because of their belief that they would lose relative to other hospitals. The teaching hospitals argued that within DRGs the patients who are treated in their institutions are more seriously ill and costly than their counterparts in other types of hospital. There is undoubtedly some substance to these arguments. However, independently of selection issues, the higher cost, less technically efficient institutions are forced under these incentives to reduce their costs to the level which is reflected in the prices received for their output.

groupings is an extension of the notion of output used in the manufacturing industry (Fetter and Freeman 1986). The total inpatient output of each hospital is defined as the number of patients treated within each predefined group. Groups are designed to be relatively homogeneous with respect to the resources used in the treatment process.

The way the English system becomes operational is through the use of (1) Diagnostic and Procedure codes, (2) Healthcare Resource Groups and, (3) Reference Costs. Diagnostic and Procedure codes – such as the International Statistical Classification of Diseases and Related Health Problems (ICD) and the Operating Procedure Code Supplement (OPCS) – are used by hospitals to capture information on diagnoses and treatments of individual patients. These codes are then used in the PbR system for allocating episodes into HRGs for costing purposes. Both the ICD and the OPCS are constantly updated, which is problematic for a researcher trying to unify subsequent versions of the same code. As of 2012, the current version of ICD is 10 and the newest revision for OPCS is 4.3. Updates and revisions usually translate into an increased numbers of codes. These are subsequently translated into more accurate tariffs. Taken together, ICD and OPCS, describe the initial diagnosis and consequent treatment of an individual patient. In order to ascribe costs to the episode of care described by disease and treatment codes, they need to be allocated to a clinically meaningful group whose resource use can be predicted to be roughly the same: the Healthcare Resource Group (HRG). This will include not only the diagnosis and associated treatment but also other predictors of length of stay and cost such as age and co-morbidities and complications.

For every Provider Hospital Spell (a period of care from admission to discharge), the patient is assigned an HRG code using a software application known as Grouper. The

Grouper (2011-2012) uses clinical coding (ICD-10 and OPCS-4) and administrative codes of patient events to derive HRGs. An HRG code consists of five characters: two letters followed by two numbers and a final letter. The first letter refers to the Chapter, the second to the Sub-Chapter; these correspond to body areas or body systems identifying the area of clinical care. Different versions have developed over time. The current version in use is version 4.0 (used in this work). This new version has been used for payments of services from 2010 onwards. Compared with the preceding version (HRG 3.5) which had 610 codes, there are now approximately 1,400 HRG groups available. This translates into a better ability to explain more of the variation in length of stay (LOS). The expanded number of codes includes not only new codes for activity in areas such as pathology but also additional degrees of complication/comorbidity and more age splits than in previous HRGs. To the extent that these improvements better target severity, they will go some way to meeting concerns about equity.

The national price per procedure defined by the HRG system implies that hospitals are not able to compete for patients on the basis of price (although hospitals are allowed to retain financial surpluses). The logic thus appears to be one of encouraging competition on the basis of quality. The danger is however that in the absence of adequate quality indicators hospitals will discharge patients early so as to shift costs to the Primary Care Trusts[7] (PCTs), increasing referral rates and reducing the use of resources in their treatment (Bevan and Robinson 2005). In the worst case scenario, the least efficient hospitals may be concentrated in areas where beds are already in short supply and the residents may be predominantly in the lower income

[7] Health care administration is an ever changing world. Until 31 May 2011 PCTs provided community health services directly. Collectively PCTs were responsible for spending around 80% of the total NHS budget. Primary Care Trusts were abolished on 31 March 2013 as part of the Health and Social Care Act 2012, with their work taken over by clinical commissioning groups.

groups (Palmer 1996). The national tariff is made sensitive to local circumstances by taking into account the unavoidable differences in the cost of providing services across the country. A Market Forces Factor (MFF) is used for this purpose by combining three separate indices - a staff index, reflecting labour market differences, a buildings index and a land index. The MFF varies in magnitude from 1.00 in West Cornwall to 1.446 for St Mary's NHS trust in London. A detailed description of the creation of annual tariffs is found in the Reference Costs Collection Guidance (Office 2008).

How do HRG compare with TFR?

The cost of each hospital episode in Clarke et al. (2003) was estimated by multiplying the length of stay by the per diem cost of the respective specialty, and the per diem costs were based on the English Department of Health's Trust Financial Returns. In this respect this costing methods is rather straightforward. Total costs under TFRs are: TC=C*d, where C= costs and d=days. C includes a number of items from overheads to the specialist's time. The shortfall however is that heterogeneity is hindered by using TFRs because the number of specialties is far inferior to the number of HRGS: 43 versus 1400. Because TFRs are averages of actual usage recorded, one does not need the full set of patient characteristics in order to obtain cost data. The use of TFRs would appear reasonable in a large and heterogeneous sample. Total costs under HRGs could instead be described as follows: TC=f(x,d)*d, where f(.) is a function that includes x= age, sex, discharge status, presence of complications or co-morbidities, hospital, operation etc (See Appendix A3.1 for complete list).

Each hospital episode in our sample was assigned an HRG4 (2011/2012) costing code using the HRG4 Grouper[8], which allocated that episode to one of 1,389 possible HRGs, organised into 23 chapters, based on body systems[9]. In our data set 19 of these chapters are represented.

The frequency of hospitalisations in each chapter is listed in Table 3.4: the most common reason for hospitalization in our sample was related to cardiac conditions. In order to have a point of reference Table 3.5 shows the distribution of HRG chapters in the English population as a whole, where the incidence of events is three times lower and cardiovascular problems are less prominent albeit still the 4th likely cause of hospitalization.

Table 3.4: Distribution of HRG chapters in the UKPDS sample 1998-2007 (N=2,791)

Chapter	Description	Freq.	Percent
E	Cardiac surgery and primary cardiac conditions	4,555	16.6
F	Digestive system	4,372	15.9
L	Urinary tract and male reproductive system	3,254	11.8
B	Eyes and periorbita	2,464	9.0
H	Musculoskeletal system	2,271	8.3
D	Respiratory system	1,943	7.1
W	Multiple trauma, emergency medicine and rehabilitation	1,450	5.3
K	Endocrine and metabolic system	1,391	5.1
A	Nervous system	1,255	4.6
S	Haematology, chemotherapy, radiotherapy	1,092	4.0
J	Skin, breast and burns	989	3.6
Q	Vascular system	967	3.5
G	Hepatobiliary and pancreatic system	600	2.2
C	Mouth head neck and ears	420	1.5
M	Female reproductive system and assisted reproduction	297	1.1
U	Undefined groups	127	0.5
V	Multiple trauma, emergency and urgent	39	0.1
R	Radiology and nuclear medicine	2	0.0
N	Obstetrics	1	0.0
Total		27,489	100

[8] http://www.ic.nhs.uk/casemix

[9] The blueprint of an HRG is as follows:

a) The first letter represents the chapter
b) The second letter represents the sub-chapter
c) The number represents the intervention or diagnosis. These four characters give the HRG root
d) The last letter represents the split for age, complications and comorbidities (CC) or length of stay (Z is used where there is no split).

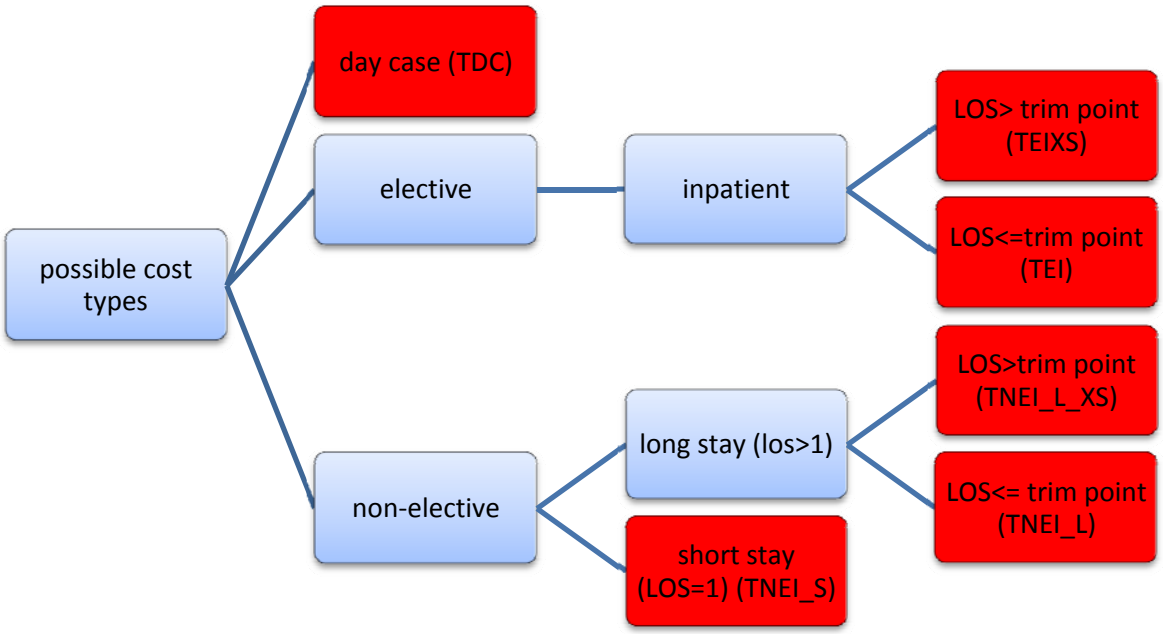
Table 3.5: Distribution of HRG chapters in the general population

Description	Total cases	% of hospitalizations	population incidence*
Digestive system	2,643,877	14.9%	5.0%
Musculoskeletal system	1,769,543	10.0%	3.3%
Diseases of childhood and neonates	1,665,220	9.4%	3.1%
Cardiac surgery and primary cardiac conditions	1,380,663	7.8%	2.6%
Obstetrics	1,359,717	7.7%	2.6%
Urinary tract and male reproductive system	1,328,419	7.5%	2.5%
Respiratory system	1,135,087	6.4%	2.1%
Immunology, infectious diseases and other contacts with health services	972,045	5.5%	1.8%
Nervous system	857,935	4.8%	1.6%
Skin, breast and burns	779,668	4.4%	1.5%
Mouth head neck and ears	764,648	4.3%	1.4%
Eyes and periorbita	627,379	3.5%	1.2%
Female reproductive system and assisted reproduction	579,194	3.3%	1.1%
Haematology, chemotherapy, radiotherapy and specialist palliative care	538,946	3.0%	1.0%
Hepatobiliary and pancreatic system	357,331	2.0%	0.7%
Critical care and high cost drug	260,213	1.5%	0.5%
Vascular system	252,194	1.4%	0.5%
Endocrine and metabolic system	224,203	1.3%	0.4%
Undefined groups	156,061	0.9%	0.3%
Multiple trauma, emergency medicine and rehabilitation	62,389	0.4%	0.1%
Radiology and nuclear medicine	4,701	0.0%	0.0%
Total	17,719,433	100%	

***** (Total cases/England population)x100. Population census 2011=53,013,000. Source: hesonline.nhs.uk

The allocated healthcare resource group was then linked to the 2012 National Health Service reference costs. Once HRGs (HRG4 2011/12) are obtained[10], reference costs can be attached to either one of the following 6 categories (Figure 3.2), depending upon whether the episodes were elective admissions, emergency admissions or day cases and whether or not these episodes are defined as short or long hospital stays:

Figure 3.2: Six possible cost categories for any given HRG



LOS=length of stay
Trim points are defined for length of stay outliers in each HRG according to whether the patient was admitted as an elective or non-elective. Excess bed days are only calculated when spell duration exceeds the trim point.

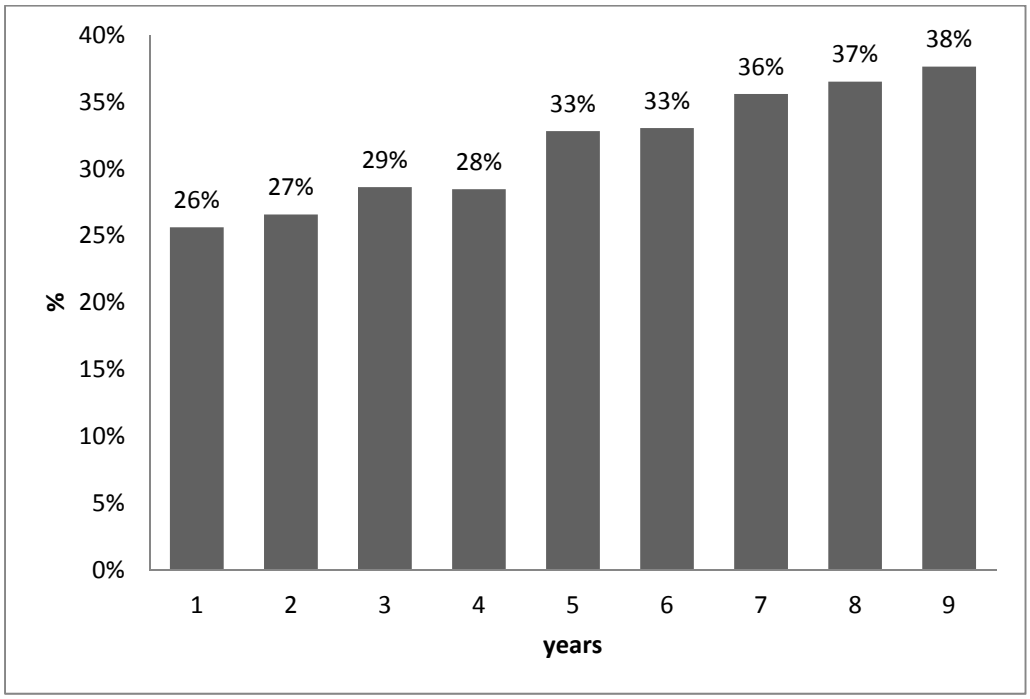
Cost code	description
TEI	Elective Inpatient HRG Data
TEIXS	Elective Inpatient Excess Bed Day HRG Data
TNEI_L	Non-Elective Inpatient (Long Stay) HRG Data
TNEI_L_XS	Non-Elective Inpatient (Long Stay) Excess Bed Day HRG Data
TNEI_S	Non-Elective Inpatient (Short Stay) HRG Data
TDC	Day Cases HRG Data

[10] With HRG4 a small number of HRGs are "unboundled". Unbundling is the first step in the grouping process, following data validation. Unbundled procedures are removed from the data input file and processed separately to derive costs. In my dataset this would include items such as renal dialysis and critical care.

Probability of incurring a hospital day case or inpatient episode and costs

Hospitalisations – defined as inpatient day cases or inpatient episodes of care – were aggregated by patient year, where years are defined as 12 months intervals from 1st October 1998 onwards. Approximately 30% of patients are hospitalized in any given year, rising from 26% in the first year to 38% in the final year (Figure 3.3).

Figure 3.3: Probability of incurring a hospital day case or inpatient episode



Conditional on having a diabetes-related complication individuals in the sample have a very high probability of being hospitalized in the same year (89%) as shown in Figure 3.4.

It is important to note that the reasons for hospitalization may not be closely related to the occurrence of diabetes complications, with 86.36% of recorded hospitalizations (5,540/6,415 patient years of HES data) occurring during years when no diabetes-related complication occurred (Table 3.6).

Figure 3.4: Probability of being hospitalized in the year of diabetes-related complication

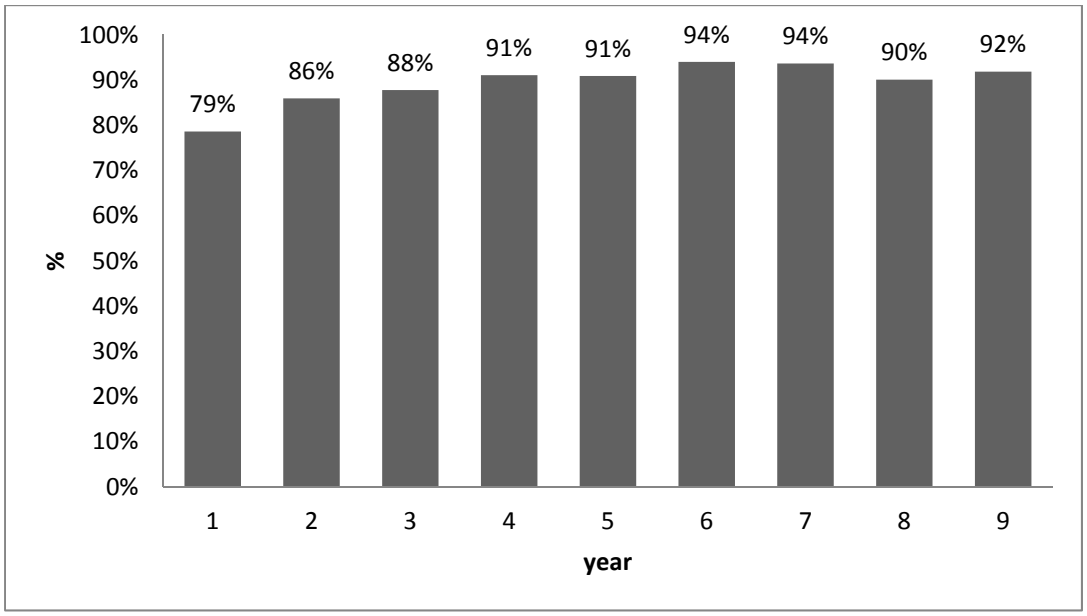


Table 3.6: Proportion of hospitalisations in years of diabetes-related complications over time.

Year	1	2	3	4	5	6	7	8	9	All
N	788	787	812	749	735	653	655	634	602	6,415
diabetes-related	103	98	108	102	110	110	74	91	79	875
%	13.1	12.5	13.3	13.6	15.0	16.8	11.3	14.4	13.1	13.6

Across all follow up periods the average annual cost conditional on hospitalization was £4,400 (SD: £8,959) and the median was £1,197 among the 2,791 patients hospitalized. The mean annual inpatient hospital expenditure was £1,306, regardless of hospitalization status; the median was £0 (Table 3.7). The predominant characteristic here is right-tailed skewness (=8.6). 80% of annual periods are without costs or contain costs of less than £5,000.

Table 3.7: Distribution of inpatient cost across years

Year	1	2	3	4	5	6	7	8	9	Total
Conditional on hospitalization (N=2,791)										
N	788	787	812	749	735	653	655	634	602	6,415
mean	4,209.98	4,497.02	4,211.29	4,295.46	4,318.99	4,219.88	4,074.17	4,970.76	4,963.97	4,400.92
SD	8,334.15	9,268.45	8,157.49	8,479.87	7,913.62	9,393.84	8,741.91	10,230.30	10,367.16	8,959.17
median	1,061.03	1,238.36	1,113.46	1,197.85	1,247.45	1,121.86	1,218.61	1,334.82	1,114.27	1,190.63
min	175.73	133.42	175.73	194.29	140.85	173.87	207.49	150.01	150.01	133.42
max	102,584.80	97,120.51	87,422.14	76,033.47	60,204.96	120,534.10	130,943.30	100,916.80	92,293.46	130,943.30
Unconditional costs (N=3,074)										
N	3,074	2,959	2,836	2,630	2,239	1,976	1,840	1,736	1,599	20,889
mean	1,079.20	1,196.07	1,205.77	1,223.31	1,417.80	1,394.53	1,450.32	1,815.36	1,868.86	1,351.52
SD	4,600.89	5,174.55	4,760.40	4,921.27	4,965.33	5,750.98	5,566.41	6,626.88	6,797.78	5,363.65
median	0	0	0	0	0	0	0	0	0	0
min	0	0	0	0	0	0	0	0	0	0
max	102,584.80	97,120.51	87,422.14	76,033.47	60,204.96	120,534.10	130,943.30	100,916.80	92,293.46	130,943.30
people never experiencing a complications (ever) (N=1,883)										
mean	607.15	629.84	738.52	707.49	708.24	802.06	903.44	928.28	1154.59	767
SE	63.85	70.41	80.7	79.44	77.05	102.19	96.91	103.61	154.87	29.47
People experiencing complications (N=1,891)										
Years with current complication										
mean	1,939.07	2,356.97	2,184.16	2,537.32	2,885.31	2,757.39	2,742.52	3,816.47	3,932.99	2,643.23
SE	235.44	284.1	252.99	294.26	305.75	387.49	426.39	473.25	555.44	110.81
Years without current complication										
mean	1,564.48	1,556.43	1,518.34	1,062.73	1,615.31	1,257.42	1,501.31	2,348.23	1,931.29	1,562.17
SE	295.44	308.07	270.78	228.41	302.96	288.7	238.61	656.42	412.97	111.21

All figures are in pound starling. The reference year is 2012.

Diabetes-related complications

Diabetes-related endpoints were recorded during routine (4-monthly) visits to UKPDS clinics, and adjudicated by an endpoint committee. These datasets were then combined to estimate the impact of diabetes-related complications on inpatient resource utilization and costs.

The average age of patients at the beginning of the follow up period was 64.3 years (SD: 9.3), and 58.3% of the sample were males. Table 3.8 shows the total number of patients incurring at least one fatal or non fatal diabetes-related complication, conditional on the patient being alive and consenting to participate in the follow-up study.

Table 3.8: **Number of complications in 3141 patients (regardless of their hospitalization status)**

Event*	Events occurring during:	
	1998-2007	Prior to 1998
MI (non fatal)	119	182
IHD (non fatal)	143	182
Stroke (non fatal)	100	96
Heart Failure (non fatal)	115	73
MI (fatal)	156	
IHD (fatal)	123	
CHF (fatal)	81	
Amputation	67	27
Blindness	70	108

*these are the number of events used for estimation. Fatal events pre 1998 are thus not accounted in costing the history of complications.

The data: Resource use questionnaire for non-inpatient care

To capture information on healthcare use other than inpatient stays and day cases, a maximum of seven resource use questionnaires were administered to patients in the trial: clinic attendees were given the questionnaires during clinic visits, and those not attending were sent a questionnaire by post corresponding to the anniversary of their therapy decision, with the exception of the first questionnaire administered between 1996 and 1997. The survey recorded information on all home, clinic and telephone contacts with general practitioners, nurses, podiatrists, opticians and dieticians and with eye and other hospital outpatient clinics. A copy of the first questionnaire is shown in Appendix A3.2.

The only difference across questionnaires was the recall period. The first questionnaire asked questions relating to use of health services over the previous *4 months*. Questionnaires 2 through 6 asked about utilization in the *previous year*. In the case of the last questionnaire (questionnaire 7), administered to all participants on the first of October 2007, the recall period was computer generated and entered into the form so as to correspond to the difference between the end of the follow up period and the date of the questionnaire last sent. The recall period thus varied between one month and one year. In the case in which the patient was alive but did not complete the 6^{th} or earlier consecutive questionnaires, the recall period was truncated and made equal to one year. Whenever at least one non-empty field in the resource use questionnaire was returned, all other fields were treated as zeros rather than missing.

On average 3.8 questionnaires were answered by 3,589 patients. The response rate was close to 90% for questionnaire 1 (Clarke, Gray et al. 2003). The response rate for questionnaires 2 to 7 ranged from 70% to 76% -- these are reported by location in

Appendix A3.3. Unlike hospitalizations, where we deal exclusively with English data, all patients participating (from England, Scotland and Northern Ireland) in the post trial monitoring (PTM) are included in the analysis.

Recruiting Nations	Number of patients	Number of complete questionnaires returned
England	2693	8411
Scotland	576	1648
Northern Ireland	320	861
Total	3589	10920

Both the QoL (see Chapter 2) and the resource use questionnaires were mailed together to participants yet we have 209 more responders (3589 instead of 3380) to the resource use than to the QoL questionnaire. However, the percentage of patients that only returned one questionnaire is higher in the resource use questionnaire than in the HRQoL questionnaire: 45% versus 42%. Table 3.9 shows the 25 most common patterns of response among patients who returned at least one questionnaire, where 1 represents an answered questionnaire and a dot indicates that the questionnaire was unreturned.

Given the different recall period across patients for questionnaire seven, patient's understanding of the recall period was examined ex-post to see if the reported resource usage was in accordance to the recall period they were given. The reported resource utilization is much higher, after annualization, than any previous questionnaire. There is not only a higher variance in costs than in any other preceding questionnaire but also averages (and medians) do not trend upwards as the recall period increases and as one would expect a priori, suggesting that most respondents were not aware of the change in the wording of the question and mechanically completed the questionnaire, assuming a one year recall period as in the previous 5 questionnaires received. These

findings are reported in appendix A3.4 as reasons for eliminating questionnaire seven from the estimation sample.

Table 3.9: Pattern of responses among participants that submit at least one complete questionnaire

Freq.	Percent	Pattern
1606	44.75	**1.....**
735	20.48	**1111111**
232	6.46	**11111.1**
101	2.81	**11.....**
81	2.26	**111....**
81	2.26	**11111..**
80	2.23	**1111...**
61	1.7	**1.11111**
53	1.48	**111111**
34	0.95	**1.1....**
33	0.92	**.111111**
24	0.67	**11.1111**
23	0.64	**1..1...**
22	0.61	**1111.11**
20	0.56	**1111..1**
19	0.53	**111.111**
18	0.5	**1.1111.**
17	0.47	**1.11...**
16	0.45	**11.1...**
14	0.39	**1.....1**
14	0.39	**1.111..**
14	0.39	**1.111.1**
13	0.36	**1..1111**
12	0.33	**111.1..**
11	0.31	**1.11.11**
255	7.11	(other patterns)
3589	100	XXXXXXX

Resource Utilization

In order to make the results of the questionnaires comparable, the responses from the first questionnaire were annualized. Table 3.10 presents information on the mean number of times patients had been seen at home by, visited, or telephoned for consultation their GP, nurse, health visitor, dietician, chiropodist, optician, hospital eye clinic, or any other hospital clinician in any given year.

Table 3.10: Resource use across all questionnaires

Variable	Mean	Std. Dev.	Min	Max
GP-Home	0.394	1.983	0	90
GP-Surgery	4.143	5.233	0	144
GP-Telephone	0.556	3.373	0	200
Nurse-Home	3.439	28.574	0	730
Nurse-Surgery	2.504	7.903	0	365
Nurse- Telephone	0.300	4.000	0	365
Health Visitor-Home	0.111	1.467	0	100
Health Visitor-Surgery	0.048	1.187	0	104
Health Visitor-Tel	0.029	0.649	0	52
Dietician-Home	0.020	0.284	0	13
Dietician-Surgery	0.225	0.956	0	30
Dietician- Telephone	0.028	0.381	0	15
Chiropodist-Home	0.305	2.099	0	104
Chiropodist-Surgery	2.037	4.627	0	135
Chiropodist- Telephone	0.063	0.668	0	33
Optician-Surgery	0.869	1.270	0	50
Optician- Telephone	0.033	0.511	0	33
Eye Clinic	1.010	1.712	0	52
Eye Clinic Telephone	0.037	0.355	0	15
Any other clinic	1.506	4.470	0	156
Any other clinic Telephone	0.081	0.683	0	33

A GP consultation at the surgery occurred 4.1 times per annum on average. This is approximately in line with the average number of consultations reported for the general population in the General Household Survey[11], where the average British patient is reported to have had 3.9 consultations each year in 1995/1996 rising to 5.5

[11]http://www.ic.nhs.uk/webfiles/publications/gp/Trends_in_Consultation_Rates_in_General_Practice_1995_96_to_2008_09.pdf

consultations each year by 2008/2009. Among the outliers, the maximum number of consultations occurred between patients and nurses. While there were no significant differences in resource utilization between questionnaires two through six, outpatient consultations with nurses, dieticians and chiropodist increased more than threefold between the first and subsequent questionnaires (Table 3.11) . This has to be set in context since the absolute rates of both telephone and visit consultations for these types of practitioners were low to start with compared with GP surgery consultations. These increases cannot be explained solely by an ageing population needing greater care and resources, but do reflect a national trend. According to the NHS information centre, the proportion of patients seen by nurses in primary care has changed over time. In 1995/1996, 75% of consultations were undertaken by GPs, 21% by nurses and 3% by other clinicians. In 2008/2009, approximately 62% of consultations were undertaken by GPs, 34% were undertaken by nurses and 4% by other clinicians. Despite all home consultations increasing, GP home consultations have remained unchanged over the years. This may also be the product of a policy change, reflecting the removal of out of hours care from GP contractual responsibilities.

Table 3.11: Resource utilization over time

	1996/7			2002-2007			
Variable	**Mean**	**Std. Dev.**	**Max**	**Mean**	**Std. Dev.**	**Max**	**P-value for equality of means <0.05**
GP-Home	0.393	2.479	90	0.394	1.701	56	
GP-Surgery	3.749	6.030	144	4.328	4.801	104	*
GP-Tel	0.426	2.922	99	0.617	3.564	200	*
Nurse-Home	1.299	14.367	297	4.447	33.174	730	*
Nurse-Surgery	1.736	7.888	294	2.866	7.885	365	*
Nurse-Tel	0.146	2.635	132	0.372	4.500	365	*
Health Visitor-Home	0.062	0.897	36	0.134	1.669	100	*
Health Visitor-Surgery	0.021	0.336	12	0.062	1.421	104	*
Health Visitor-Tel	0.009	0.209	6	0.039	0.774	52	*
Dietician-Home	0.005	0.143	6	0.027	0.330	13	*
Dietician-Surgery	0.167	0.801	15	0.252	1.019	30	*
Dietician-Tel	0.011	0.232	9	0.036	0.433	15	*
Chiropodist-Home	0.083	0.834	27	0.409	2.473	104	*
Chiropodist-Surgery	1.289	2.887	48	2.390	5.213	135	*
Chiropodist-Tel	0.061	0.738	33	0.065	0.633	23	
Optician-Surgery	0.666	1.559	33	0.965	1.095	50	*
Optician-Tel	0.045	0.840	33	0.026	0.228	6	
Eye Clinic	0.672	1.675	27	1.170	1.706	52	*
Eye Clinic Tel	0.027	0.372	12	0.041	0.347	15	
Any other clinic	0.991	2.720	42	1.749	5.072	156	*
Any other clinic Tel	0.057	0.787	33	0.093	0.627	22	*

* stands for `yes'

Total non-inpatient Costs

To calculate the total costs of these services, unit costs were obtained for each category of health service use, and these are presented in Appendix A3.5, alongside information on the sources of the values used. These unit costs were then applied to the quantities of resource use to obtain costs per category of health service use. Total costs were then obtained by multiplying the number of contacts by unit costs for each type of resource use and expressed on an annual basis. The baseline year for pricing used is 2012.

On average conditional on returning a questionnaire, 6% of individuals incurred no out of hospital expenditure in any given year. However this masks great differences of utilization across questionnaires as shown in Figure 3.5.

Average annual non-inpatient expenditures vary greatly across individuals, with a standard deviation of £1081.035, which is more than 1.5 times the mean £676.2, across all questionnaires. The median is considerably smaller than the mean, reflecting the skewness of the data. Kurtosis indicates that the tails are much thicker than those of a normally distributed data (88.36). For readability, the very long right tail of the distribution of costs in Figure 3.6 has been truncated at £5,000. When costs are re-expressed on a logarithmic scale, the distribution of costs appears to approximate well that of a normal distribution. However in the logarithmic transformation only strictly positive expenditures are considered, it thus ignores the 259 individuals (out of 3589) who returned questionnaires with no resource utilization.

Figure 3.5: Proportion of patients with no resource utilization per questionnaire round

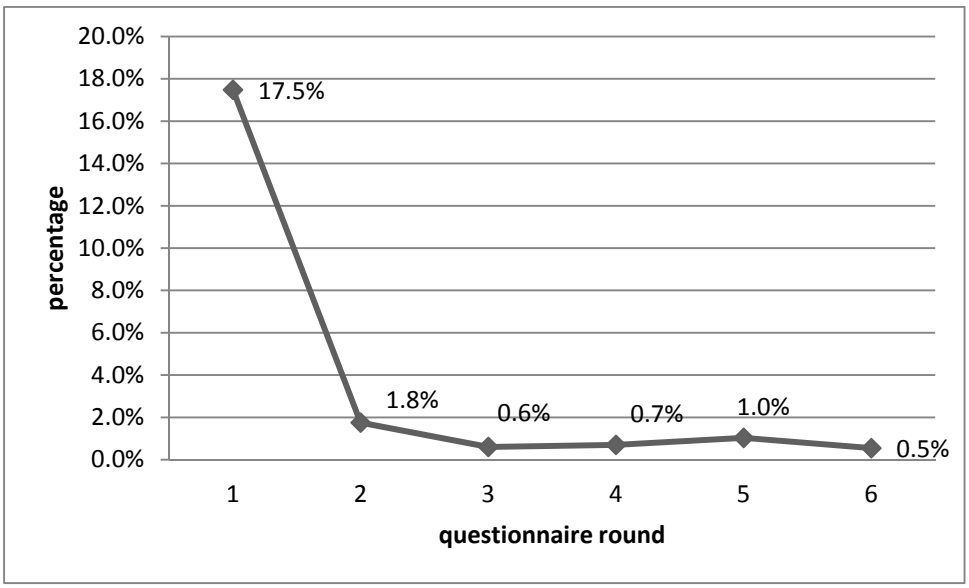
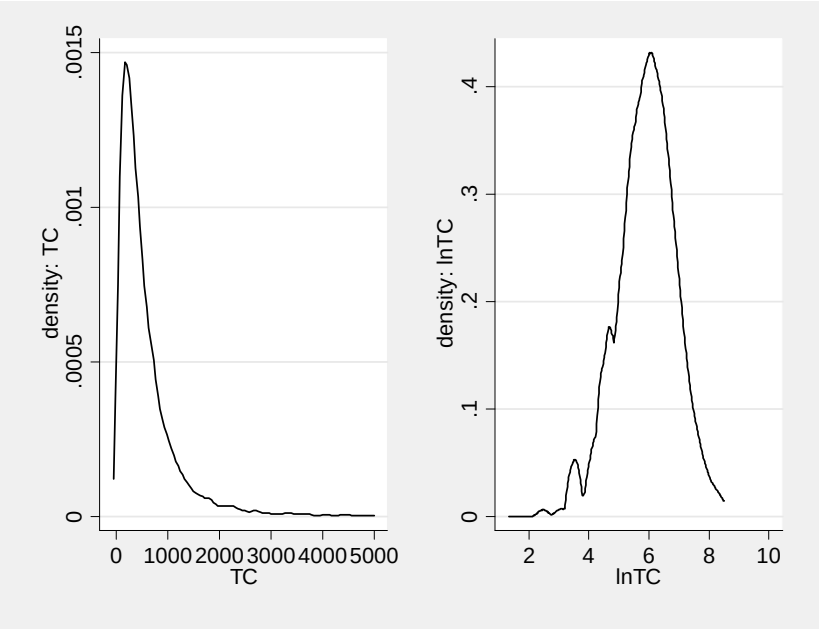


Table 3.12: Total average annual non-inpatient cost per questionnaire round

Costs	Q1	Q2	Q3	Q4	Q5	Q6	Total
N	3495	1655	1654	1566	1453	1097	10920
Unconditional costs (N=3,589)							
Mean	479.22	746.79	755.17	740.56	818.73	797.67	676.21
Std. Dev.	787.16	1195.80	1093.35	1055.50	1394.11	1162.90	1081.04
Median	267.51	476.00	484.10	490.60	489.10	515.00	431.10
Skewness	7.56	7.73	6.62	7.62	7.14	5.98	7.48
Kurtosis	112.13	86.96	69.24	97.31	72.46	48.62	88.36
People never experiencing a complications (N=2,271)							
mean	426.43	600.49	606.02	597.60	677.35	630.06	550.75
SE	16.56	28.54	25.30	25.77	45.44	37.34	11.11
People experiencing complications (N=1,318)							
Years with current complication							
mean	593.05	1,034.20	980.67	1,017.72	1,109.75	1,045.05	937.53
SE	40.52	96.09	78.94	81.16	101.59	85.55	32.65
Years without current complication							
mean	557.30	807.04	876.50	796.06	858.74	924.48	757.76
SE	25.89	52.86	60.80	54.08	59.90	85.96	20.66

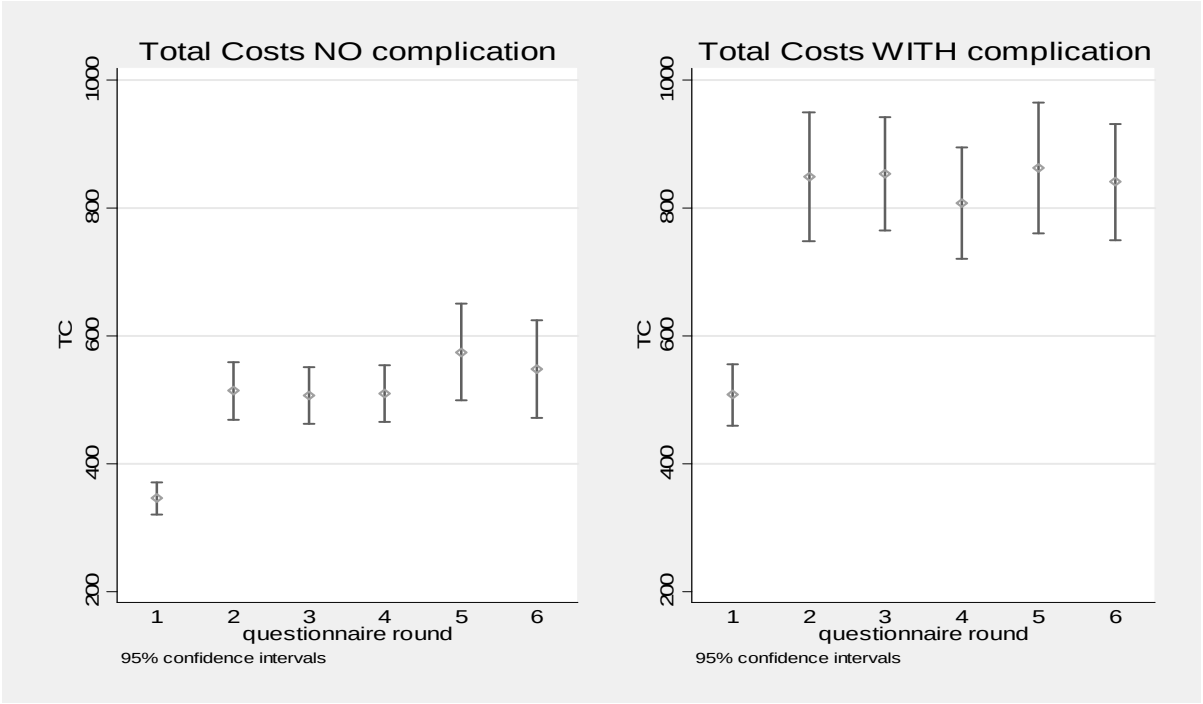
Figure 3.6: Distribution of costs in levels and in logarithms



Given that questionnaire one had a different recall period to the subsequent questionnaires used and that, even after annualization, the mean cost was considerably

lower than in the questionnaires administered between 2002 and 2007, an important question to ask is whether it is justified to include the first questionnaire for analysis or not. One way of answering this is to examine the proportional effect of diabetes related complications in each questionnaire round, and Figure 3.7 shows that these effects are of similar magnitude. Thus, once properly accounted for by the use of an indicator in a model specified on a multiplicative scale, the first questionnaire can be safely analyzed together with all subsequent questionnaires (Appendix A2.6 analyses in more detail the relationship between recall window and accurate extrapolation).

Figure 3.7: Annual costs with and without complications per questionnaire round



Complications

Table 3.13 shows the total number of complications incurred by patients answering at least one resource use questionnaire and the number of patients experiencing them. Amputations and cataract extractions are more likely to reoccur than any other complication.

Table 3.13: Total number of events and number of responders experiencing them

Non Fatal Events	Number of events	Patients with events
MI	451	384
IHD	435	428
Stroke	275	246
Amputation	163	108
Blindness to 1 eye	248	227
Heart Failure	196	193
Cataract Extraction	1066	712

The upper triangular matrix in Table 3.14 shows the pairwise correlations at the 5% significant level. The diagonal values represent the number of patients that experienced a certain event and the lower triangular matrix represent the number of patients having at least one of any two types of complications, for example, 65 patients have both an MI and an IHD, 33 patients have both an MI and a stroke, and so on.

Table 3.14: Correlation matrix and participants with one or more complication answering at least one questionnaire

Number of patients with one or more events/ ***correlation matrix***	MI	IHD	Stroke	HF	Amp	Blind	Cat
Myocardial Infarction	**384**	0.047*	0.028*	-0.0001	0.013	0.015	0.014
Ischaemic Heart Disease	65	**428**	0.031*	0.002	0.030*	0.014	0.027*
Stroke	33	41	**246**	0.054*	0.036*	0.028*	0.039*
Amputation	13	17	16	**108**	0.019*	0.055*	0.030*
Blindness in 1 eye	24	34	20	14	**227**	0.032*	0.030*
Heart Failure	15	17	21	11	21	**193**	0.235*
Cataract Extraction	85	104	50	34	122	49	**712**

*P-value <0.5

In order to test for the presence of multiplicative effects, the following specification is tested:

$$E(TC_{it}|\boldsymbol{X}_{it}) = \alpha + \beta \boldsymbol{X}_{it} + \gamma \boldsymbol{X}_{it}\boldsymbol{X}'_{it}$$

where **X** is the vector of complications and the joint hypothesis that γ=0 is tested. For 7 complications, there are 21 possible correlation to be tested ((7*6)/2), and jointly one cannot reject the hypothesis of independence (F(21, 10899) = 3.30 Prob > F = 0.001).

The data reveal that some patients who report never using non-inpatient services do have complications and, that those patients without complications prior to a questionnaire do nevertheless frequently see doctors and other practitioners outside the hospital setting. Approximately 21% (55/259) of those who never reported any non-inpatient health care use do have complications. Similarly, as illustrated in Figure 3.7, conditional on never having a complication, the majority of patients (90%) still use non-inpatient medical services.

Table 3.15: Patients' complications and costs

Matrix	TC=0	TC>0	**Total**
No Complication	204	1,784	1,988
Complication	55	1,546	1,601
Total	259	3,330	**3,589**

Patients experiencing macro vascular complications are less likely to report higher non-inpatient utilization compared to those that have micro vascular complications. This could be due to the following: 1) individuals who experience a macro-vascular event tend to have more fatal events in the study period and a shorter life expectancy than those who experience a micro-vascular event. 2) a higher proportion of macro-vascular events are treated in hospitals and so this might be associated with lower outpatient resource utilization than micro vascular events.

The pattern of resource use is higher (two or three times more per year) among those who experienced a complication than among those who never had a complication. It is important to recognize however that there may be two effects at play: on one hand resource utilization is likely to be higher conditional on having a complication (as Table 3.16 shows), but on the other hand, resource utilization may serve a preventive role and thus reduce the incidence of complications.

Table 3.16: Annual resource use by type of service among patients with and without complications

Resource	No Complication (N=1,988)	sd	With Complication (N=1,601)	sd	Difference in utilization with vs. without complications
GP-Home	0.63	3.01	1.19	3.30	0.56 **
GP-Surgery	5.49	6.69	6.95	7.86	1.45 **
GP-Tel	0.90	5.75	1.45	3.59	0.56 **
Nurse-Home	3.31	26.43	12.20	54.56	8.89 **
Nurse-Surgery	3.42	10.21	5.35	16.01	1.92 **
Nurse-Tel	0.44	2.24	1.08	10.04	0.64 *
Health Visitor-Home	0.18	1.97	0.43	3.06	0.25 **
Health Visitor-Surgery	0.11	2.39	0.17	1.57	0.06 *
Health Visitor-Tel	0.03	0.29	0.14	1.62	0.11 **
Dietician-Home	0.02	0.23	0.10	0.68	0.08 **
Dietician-Surgery	0.34	1.02	0.67	1.88	0.33 **
Dietician-Tel	0.05	0.50	0.09	0.73	0.04 *
Chiropodist-Home	0.38	2.73	0.92	4.01	0.54 **
Chiropodist-Surgery	2.28	5.89	3.76	7.31	1.48 **
Chiropodist-Tel	0.11	0.80	0.22	1.38	0.11 **
Optician-Surgery	1.10	1.68	1.57	2.11	0.47 **
Optician-Tel	0.07	0.82	0.11	0.95	0.04 *
Eye Clinic	0.84	1.38	2.50	3.04	1.66 **
Eye Clinic Tel	0.04	0.42	0.16	0.75	0.11 **
Any other clinic	1.96	3.99	3.75	7.94	1.79 **
Any other clinic Tel	0.17	1.11	0.28	1.19	0.11 **

P-value<0.01 **, P-value<0.05 *.

3.4 Methods

General Approaches to modelling costs

There are a number of data issues one needs to consider before attempting estimation. Health care costs are typically positively skewed because a small number of patients incur particularly high costs while it is not possible to have negative costs. We usually also face a substantial fraction of zeros because, as seen above, many individuals are not hospitalized in any period even if they experience complications and likewise some individuals do not incur any non-inpatient costs.

Mihaylova and co-authors (2010) reviewed a number of statistical methods for analysing healthcare resource use and cost data, their ability to address skewness, multimodality and heavy right tails, and their ease for general use. The review identified the following categories: (1) methods based on the normal distribution, (2) models based on normality following a transformation of the data, (3) single-distribution generalized linear models (GLMs), (4) parametric models based on skewed distributions outside the GLM family, (5) models based on mixtures of parametric distributions, (6) two-part models, (7) Heckman selection and Tobit models, and (8) semi-parametric and non-parametric methods. Based on this review, their recommendations are that, first, simple methods are preferred in large sample sizes (in the thousands) where the near-normality of sample means is assured. Second, in somewhat smaller sample sizes (in the hundreds), relatively simple methods, able to deal with one or two of the data characteristics studied, may be preferable, but checking sensitivity of results to assumptions is necessary.

The choice between a two part model (6) and a generalized Tobit or sample selection specification (7) to model the demand for medical care deserves to be outlined first as it has been a contentious topic in the literature for the past three decades. The two-part model is an attempt to deal with the problem of zero-cost observations by separating behaviour into two stages. Stage one is any cost incurred, and stage two: conditional on some cost being incurred, its level. More formally, the model has two equations. The first could be a probit, logit or linear probability model for the dichotomous event of having either zero or positive expenses. Log linear OLS, GLM or GEE models have been employed in the second part to evaluate mean costs, and truncated Poisson-lognormal models to evaluate resource use (Winkelmann 2004). The crucial assumption here is that the two parts are independent. This means that the explanatory variables in the conditional cost equation are independent from the error term in the selection equation and that the error terms in both equations are independent. This is not too restrictive if we can control for those variables in the selection equation that are correlated with the error term in the equation of interest. However, one need not impose independence among the error terms in the two equations. The two alternative models that allow for this type of correlation are the Tobit model, and the sample-selection or Heckman model. The Tobit model assumes that the dependent variable follows a censored normal distribution where the censoring function (e.g. whether an individual has incurred a cost or not) and the expense function (e.g. how much?) have the same coefficients. One caveat to bear in mind is that although the Tobit model makes the same assumptions about error distributions as the OLS model, it is much more vulnerable to violations of those assumptions (even if bootstrap or jackknife standard errors are computed). The sample-selection or Heckman

model, allows the censoring function and the expenditure function to have different coefficients and the error terms to be correlated.

Duan et al. (1983) reject the Tobit model in favour of a split sample analysis for 3 reasons: First, Tobit assumes that the functional form is known a priori and that the functional form must be such as to yield a bivariate normal error. Second, they argue, a Tobit model is equivalent to a sample-selection model whose interpretation is inappropriate for modelling health expenditure data because it estimates an unconditional (uncensored) expenditure equation that describes the expenses that all individuals (including non-spenders) would have if they were all spenders. However, they argue, one ought not to be interested in such an equation because the zero spenders are not cases with missing expenses but rather individuals who incurred no costs. Third, the self-selection model has poor numerical and statistical properties. The likelihood function of the selection model is known to have non-unique local maxima. In contrast, a two-part model has a unique global maximum. The critiques of the Duan paper arise in the context of a demand model for health insurance. The debate was refuelled following the use of two part models in the RAND Health Insurance Experiment. Tooze et al. (2002) argued then for a selection model that explicitly estimates the correlation between a probit and lognormal model for repeated measures data.

Manning et al. (1987) use Monte Carlo simulations to compare the Heckman Two Stage model (2SM) and the two-part model (2PM) full maximum likelihood estimator (FIML). They find that 2PM outperforms 2SM by comparing these in terms of mean squared forecast error. Collinearity problems and violation of the bivariate normality assumption for the error term led to poor performance of selectivity models. However a

limitation of the standard 2PM is that it is based on the assumption of independence between part one and two – estimates would be inconsistent if this assumption is violated. On the other hand, when we consider 2SM, empirical results are sensitive to misspecification if the normality assumption of the error term in both equations does not hold[12].

Models based on mixtures of parametric models (5) are introduced as a flexible way to accommodate excess zeros, over dispersion and heavy tails by allowing for different parts of the response distribution to be differently affected by covariates i.e., rare users, low cost users and high cost users. The two-part model is a special case of a mixture model with only two components one of which is degenerate. Two-part models are shown to perform better than single-equations models as they accommodate heterogeneity between users and non-users (Duan, Manning et al. 1983).

Conditional on hospitalization and thus when dealing with data without zeros, there are two major classes of alternatives to the traditional least squares formulation used to address the data issues endemic to cost data conditional on expenditure. One strategy seeks to transform the data into logarithms using MLE or Box-Cox transformations. Often the OLS (ordinary least squares) approach is used following transformation of data. Because ultimately interest lies in the original scale, a back transformation of the results to the original scale is required. Different approaches for back transformation are available according to the nature of the error term. Under the assumption of normality of the error term, the mean of the untransformed scale can be directly estimated from the estimates of model parameters and variance. However,

[12] One should note that limited information maximum likelihood (LIML) also known as 2-step estimator does not assume a bivariate normal distribution as FIML, it is also faster to compute and thus more stable.

reliance on the assumption of normality or homoskedasticity can lead to inconsistent estimates. For back transformation one should use Duan's non parametric smearing when the error term is homoskedastic (Duan 1983) and its variants when the error term is heteroskedastic and the controls of interest are indicator variables (Manning 1998). The methods employing an initial data transformation are shown to provide potentially more efficient estimates in heavy tailed data but can perform badly if an inappropriate transformation is used (Briggs, Nixon et al. 2005).

To avoid the potential problems in retransformation, researchers often opt to allow the linear model to be related to the response variable via a link function and for the magnitude of the variance of each measurement to be a function of its predicted value. In this way the retransformation problem is eliminated (Manning and Mullahy 1990). There are however many link functions, and their choice can be somewhat arbitrary. The most common approach for choosing the appropriate link and family functions has been to rely on a series of diagnostic tests for candidate link and variance function models. It is important to note however that, if the link function is wrong, the results will be biased, but if the family is wrong, then there will be losses in efficiency but results will not be biased (Manning and Mullahy 2001). Finally, one should keep an eye on the interpretation of different estimators. In ordinary least squares (OLS) the effect of a covariate on the outcome is additive; in log transformed OLS and in GLM with log link the effect is multiplicative. The most widely used GLM model for modelling cost in the literature uses the gamma family with log link. Generalized estimation equations (GEE) are also used (Basu and Rathouz 2005) with a similar specification to estimate population average effects and unlike GLM models, they are less sensitive to the specification of the variance structure.

Basu and Manning (2009) while discussing the strengths and limitations of existing methods concluded that no current method is optimal or dominant for all cost applications. Many of the diagnostics used in choosing among alternatives have limitations that need more careful study. Several avenues in modelling cost data remain unexplored.

Causality

Much of the literature treats complications as completely random processes in costing models as illustrated by the prevalence of population average and pooling models. This exogeneity assumption is plausible if in the absence of an event, expected costs were zero, homogeneity across patients were to hold, or if there is no evidence that the likelihood of diabetes complications is correlated with other non-diabetes-related drivers of cost. However in the situation considered here, patients incur costs which may be related to non-diabetic events or to preventive-medicine activities designed to reduce the risk of complications. Therefore, the single intercept assumption (all patients having the same baseline level of costs in the absence of complications) may not hold as patients with and without complications may have different patterns and levels of utilization both before and after events.

Whether or not there is a real belief in exogeneity underlying the use of population models – the literature is more concerned about functional form issues than issues of causality. This is understandable because the techniques available to deal with causality, and specifically within patient variation, require the relationship between the regressors and the conditional mean to be linear, which might lead to bias or imprecision because the model might either be mis-specified or might fail to take advantage of properties of the distribution of the error term. In fact models that explore

multiplicative effects between covariates and outcomes to accomodate skewness are inherently non linear. A linear estimation equation may be achieved through logarithmic transformation but this is sometimes discarded in this context for two main reasons: the impossibility of incorporating zero value observations and the retransformation problem. In addition, cost benefit analyses (CBA) and budget models are primarily concerned with sample prediction and as such should be willing to do some trade off between biased estimates for causal parameters and the precision of the estimators used[13]. For example, fixed effects purge any potential endogeneity attributable to time-invariant unobservables but do so at the cost of throwing away all of the between patient variation in the data. It is for this reason that population average models are more efficient than FE models.

Table 3.17 decomposes unconditional inpatient (HES) and non-inpatient (resource use questionnaires) costs into within patient and between patient components in order to show that the potential efficiency loss from exploiting only within-patient variation is large. The between numbers refer to deviations across patients and the within numbers refers to the deviation from each individual's average. The variation in inpatient costs is slightly higher within patient over time than across patients, while the opposite is true for non-inpatient unconditional costs. This suggests

[13] ***Bias*** is the average difference between the estimator and the true value (unknown as the data generating process itself unless we simulate data). ***Precision*** is the standard deviation of the estimator. One measure of the overall variability is the Mean Squared Error, MSE, which is the average of the individual squared errors. The MSE can also be expressed as follows MSE=(precision)^2+(bias)^2. Often the overall variability of a biased estimator is *smaller* than that for an unbiased estimator, in which case the biased estimator is superior to the unbiased one.

that individuals' propensity to use medical resources does not change much over time when it comes to GP visits and other non-hospital related expenditures as opposed to hospitalizations where the majority of care is non-elective.

Table 3.17: Between and Within Decomposition of unconditional costs

	inpatient		non- inpatient	
	mean	SD	mean	SD
overall	1,351.52	5,363.65	676.2	1081.04
between		3,917.00		863.7
within		4,500.56		697.4

It is likely however that for budgetary purposes marginal differences for heterogeneous groups may be less interesting (being less precise) than population averages in expenditures. Appendix A3.10 investigates the possible presence of fixed effects in both inpatient and outpatient healthcare costs. While fixed-effects produce on average lower probabilities of hospitalizations than pooled estimates, they report higher marginal impact of complications on costs than pooled estimates. The unconditional effects are dominated by the impact of complications on costs over the probability of incurring costs and thus for the unconditional costs, we get higher expenditures under FE for patients under 60 year of age. Cost rise faster under the gamma distribution than under the normal distribution as patient's age increases. Differences in results can arise from one of two reasons: estimation or identification.

In terms of estimation, assuming the data generating process can be defined as:

$$C_{it} = \beta X_{it} + \gamma Z_i + u_i + e_{it}$$

and that there exists a time-invariant variable Z that we do not observe and which captures heterogeneity. Within-differencing eliminates the counfundeness that would arise from not observing Z_i. In order for the coefficients of the FE estimation to be

higher than those of the PA it must be the case the Z is negatively correlated with costs (C) and positively correlated with complications (X) or vice-versa. If Z is correlated with u (the time-invariant error term) by construction it is also correlated with C and the sign of the association is maintained. Taking the subsample of patients contribution to the identification of FE and thus holding constant the sample for identification, I show that results are driven by the sub-sample that helps towards the identification of within differences. In essence, FE cuts down on the sample for identification and this subsample is not representative of the UKPDS population as a whole. As events are rare, some individuals will help identify some coefficients but not all. The inclusive sample is a null set as there is no patient who helps identify all complications analyzed. I therefore test the hypothesis of results being driven by the subsample responsible. In Appendix A3.10, I show that differences between FE and pooled estimates are dictated by the sample needed for identification rather than by the estimation strategy by rejecting the hypothesis that the coefficients are the same across the pooled sample and the FE sample using the same estimation method across the two samples. I also show that in the case of inpatient costs the multiple intercept hypothesis that would call for the use of FE is not statistically significant when tested on the subsample of contribution patients. FE is present in the subsample that identifies within variation in non-inpatient costs suggesting that the level of utilization of non-inpatient resources are more constant over time and across patients than inpatient utilization.

3.5 The Model for inpatient costs

Herein the aim is to estimate the mean inpatient/day case health costs in a cohort of diabetes patients, including those individuals who have no medical expenditure during

the period of interest and to examine the relationship between inpatient healthcare costs and a number of predefined clinical events.

In order to derive the most appropriate approach for this purpose, I start by considering the nature of the variables under consideration: a binary indicator, h_i that denotes whether a person has incurred a cost or not; a set of covariates **x** and, a continuous dependent variable c_i denoting costs (representing either hospitalization costs or other resource use expenditures). We know that amongst the individuals who do make use of health services, the distribution of costs is skewed due to the presence of a small number of individuals incurring disproportionately high healthcare costs.

Following Andrew Jones's three-part taxonomy to determine the appropriate approach in this type of situation (Jones 2000), I start by stating that $C_i=0$ here represents a genuine zero, as it is safe to assume that patients have no incentive to avoid hospitalization when they need it, especially since patients enrolled in the trial are monitored at periodic intervals and have free access to medical care. In other words, we shall safely assume that there are no problems of unobservables or self-selection. We are able to observe perfectly attrition from the sample – the data set allows us to know, in addition to dates of deaths, which patients migrate and when, and so we have close to complete follow up (at three to four month intervals). Conditional on being in the realm of genuine zeros, the second question that Jones poses is whether the fact that one has incurred a cost is influenced by the decision of how much to consume. If the answer is "no", he suggests a *sequential decision model* is appropriate. If the answer is "yes", a *joint decision model* is appropriate. In the context of my research question, the answer to this question it is reasonable to assume that a patient who incurs a complication has no power in deciding the occurrence, type, timing or cost of a

diabetes-related complication (one should bear in mind however that the probability of being hospitalized could be positively correlated with the level of costs). Sequential choices are often used to motivate the *two-part model*, while joint decisions are associated with *generalized Tobit* and *selection models*. The final question concerns the object of the analysis. Is the object simply to predict E(C|x) – for example to deal with the problem of imputing missing values due to non-response to an item in a sample survey? Or is the object to make inferences about the parameters? The answers to these questions determine the appropriate method to adopt.

As a broad modelling strategy, herein I outline the preferred modelling strategy: a 2PM. This model assumes independence between the probability of hospitalization and the conditional costs and population averages.

A two-part model (2PM)

This section closely follows the methodological approach undertaken by most of the literature looking at the cost of diabetes complications and in particular it replicates the strategy adopted by previously published work. Usually in the first stage a logit or probit model is used. These two models are very similar, the only difference between these are the assumptions underlying the error term: probit assumes a normal distribution while logit assumes a logistic distribution[14] (for a comparison of these alternatives please refer to Appendix A3.6).

[14] A third option is to use a linear probability model (LPM), in other words an OLS model to estimate a binary dependent variable model. The LPM is computationally simpler and easier to interpret than both probit and logit. However, the biggest problem with a LPM is that the predicted values are not necessarily guaranteed to be constrained between the 0 – 1 interval, unless all explanatory variables are binary.

Because of its ease of interpretation a logit model with clustered standard errors is used to model the probability of incurring costs within a single patient-year time period. The dependent variable is set equal to 1 in any patient year an individual incurs costs.

Part 1-- Logit Random Effect model:

$$\Pr(C_{it} > 0 | X_{it}^{j}) = \frac{e^{\alpha_0 + \alpha_j x_{it}^{j}}}{1 + e^{\alpha_0 + \alpha_j x_{it}^{j}}}$$

$$\Pr(C_{it} = 0 | X_{it}^{j}) = 1 - \frac{e^{\alpha_j x_{it}^{j}}}{1 + e^{\alpha_j x_{it}^{j}}}$$

This has the desirable property that the log of the odds is a linear function of the explanatory variables.

$$ln\left(\frac{\Pr(C_{it} > 0 | X_{it}^{j})}{\Pr(C_{it} = 0 | X_{it}^{j})}\right) = \alpha_0 + \alpha_j x_{it}^{j}$$

Where α indicates the estimated coefficient for the j^{th} variable of interest and the subscripts i and t indicate the patient and the time period respectively.

In our data all patients with a fatal stroke are hospitalized in the year this event occurs. This means that the fatal strokes perfectly predict the outcome and thus that the coefficient of this event on the probability of hospitalization cannot be estimated in a logit (or probit) setting as the coefficient would in this case approach infinity.

In the second part the total costs incurred are estimated, conditional on hospitalization. Here to address the skewness present in the data, I use a flexible GEE model[15] (Basu and Rathouz 2005).

[15] Parameter estimates from the GEE are consistent even when the covariance structure is misspecified, under mild regularity conditions. The focus of the GEE is on estimating the average response over the population ("population-averaged" effects) rather than the regression parameters that would enable prediction of the effect of changing one or more covariates on a given individual. GEEs are usually used in conjunction with robust standard error estimates. GEEs

Part 2 – Generalized Estimation Equation (GEE)

$$g\{E(C_{it})\} = x_{it}\beta, \quad y \sim F \; with \; parameters \; \theta_{it}$$

g( ) is called the link function, and F is the distributional family. Substituting various definitions for g( ) and F results in a wide array of models. For instance, if C_{it} is distributed Gaussian (normal) and g( ) is the identity function, we have

$$E(C_{it}) = x_{it}\beta, \quad C \sim N$$

yielding a linear, random-effects regression, or other regression-related models, depending on what we assume for the correlation structure.

In order to specify the relationship between the set of regressors and the conditional mean I test for two link functions: the identity link and the log link (see Appendix A3.7). The former has the benefit of ease of interpretation, as the coefficients read in the same way as OLS irrespective of the choice of distribution. In the latter, covariates act multiplicatively on the mean, changing the interpretation of the coefficients[16] Unlike log models however, no retransformation is needed as the link function applies to $E[C_{it}]$ and not to the cost itself C_{it}. The link test cannot reject the identity link as covariates act additively on the mean.

$$E(C_{it}|C_{it} > 0, X_{it}) = \beta^j X_{it}^j$$

In order to describe the relationship between variance and conditional mean a gamma distribution is used: $var[C|x] = (E[C|x])^2$ as costs are constrained to be positive. This places less weight on outliers (i.e., patients with high costs) than the

---

are thus de facto a semi-parametric regression techniques because they rely on specification of only the first two moments. Under correct model specification and mild regularity conditions, parameter estimates from GEEs are consistent. They are a popular alternative to the likelihood–based GLM which is more sensitive to variance structure specification.

[16] This is different to log transformation of the dependent variable (another common approach with skewed data) ln(E[y|x])≠E[ln(y)|x]

normal distribution. To determine the appropriate family (distribution) function among the possible alternatives, the squared raw-scale residual in a modified Park test is used (See Appendix A3.8). The modified Park test (the actual Park test is a test for heteroskedasticity) exploits the idea that a distribution reflects the relationship between the variance and the mean and that regressing the square of the residuals on the natural log of the predicted cost would provide guidance on the appropriate distribution family (Basu and Manning 2009).

$\text{var}[y|x] = (E[y|x])^{\lambda}$

Test for λ:

λ=0: Gaussian
λ=1: Poisson (variance proportional to the mean)
λ=2: Gamma
λ=3: Inverse Gaussian

The two parts can be combined to estimate the expected unconditional expenditure by exploiting the following probability rule:

$$E(C_{it}|X_{it}) = \Pr(\mathrm{C_{it}} > 0) \times \mathrm{E}(C_{it}|C_{it} > 0)$$

The expected unconditional cost is the product of the expectation of estimates from parts 1 and 2. The above equation assumes that the error terms in each of the two models are independent.

In order to obtain a measure of uncertainty for the estimated unconditional cost, the 95% confidence intervals (CIs) are calculated through non-parametric bootstrapping. This is done by randomly re-sampling individuals with replacement and re-estimating the equations. Costs are then calculated and rank-ordered, and the 2.5th and 97.5th estimates are taken as the confidence limits. Bootstrapping of unconditional costs provides a convenient way of capturing joint uncertainty around the two equations, and

also recognises the possibility that higher probabilities of hospitalization may be associated with higher costs.

3.6Results

Hospitalization costs

Table 3.18 reports the estimated coefficients for the two-part model most often used in the literature to predict hospital in-patient costs. The first part equation is a random effect (random intercept) logit model with clustered standard errors. The slope coefficient is interpreted as the change in the log odds as the value of the explanatory variables of interest moves from 0 to 1. From this we can see that the odds of being hospitalized in the same year of a diabetes-related complication are much higher than the odds of being hospitalized conditional on having a history of diabetes-related complications. In the second equation, I use a GEE model with identity link and gamma distribution.

In order to report unconditional costs I resort to a representative patient. Using the values for the relevant coefficients in Table 3.18, the probability of hospitalization for a 60 year old male following a non fatal MI is as follows:

$$\Pr(C_i|\,\text{nf MI}{=}1) = \frac{\exp\left(-1.253 + 0.043 \times (60-65) - 0.120 \times 1 + 4.495 \times 1\right)}{1 + \exp\left(-1.253 + 0.043 \times (60-65) - 0.120 \times 1 + 4.495 \times 1\right)}$$

The total hospital costs conditional on the patient attending hospital during the same year can then be calculated using the following equation:

$$E(C_i|C_i > 0, \text{nf MI}{=}1)$$

$$= 3{,}251.621 + \ -234.628 \times (60-65) - 38.11 \times 1 + 3{,}788.652 \times 1$$

Table 3.18: Regression equations to derive annual probability of incurring hospital costs (1), and costs of hospital care conditional on incurring costs (2)

	(1) hosp		(2) cost	
	Coeff	SE	Coeff	SE
Constant	-1.253**	(0.046)	3,251.621**	(156.466)
Age-65	0.043**	(0.003)	38.110**	(9.816)
Male =1	-0.120*	(0.058)	-234.628	(196.251)
Event during year indicator				
MI	4.495**	(0.451)	3,788.652**	(1,099.558)
IHD	3.366**	(0.282)	8,198.890**	(1,638.188)
Stroke	2.394**	(0.286)	6,742.828**	(2,228.644)
HF	2.960**	(0.303)	1,011.166	(951.429)
Amputation	4.063**	(0.556)	7,678.835**	(1,607.236)
Blindness 1 eye	0.847**	(0.284)	2,452.571	(1,679.667)
Fatal MI	5.095**	(0.539)	-1,295.735**	(441.573)
Fatal IHD	4.682**	(-0.544)	969.746	(1,065.327)
Fatal Stroke			1,041.983	(1029.867)
History†				
MI	0.676**	(0.099)	1,353.651**	(447.833)
IHD	0.563**	(0.092)	1,926.805**	(420.311)
Stroke	0.434**	(0.123)	2,434.340**	(836.556)
HF	0.826**	(0.143)	1,915.375	(1,087.357)
Amputation	1.276**	(0.218)	1,595.121*	(785.079)
Blindness 1 eye	0.260*	(0.122)	-480.919	(312.757)
Observations	20,889		6,415	
Number of patients	3,074		2,389	

†Likelihood-ratio test of rho=0: chibar2(01) = 1188.19 Prob >= chibar2 = 0.000
** p<0.01, * p<0.05

Table 3.19: Estimated probability of incurring some hospital in-patient costs (1), estimated annual hospital in-patient costs conditional on cost being incurred (2), and expected mean cost (3), for a representative individual (Male, age 60) during year of complication and in subsequent years

	(1) hosp			(2) cost			(3) hosp*cost		
	Mean	95% CI Lower	95% CI Upper	Mean	95% CI L	95% CI U	Mean	95% CI L	95% CI U
no event	15.7%	14.3%	17.9%	£2,912	£2,721	£3,270	£459	£462	£600
Event during year indicator									
MI	94.4%	89.6%	98.4%	£6,756	£4,778	£8,897	£6,379	£4,500	£8,495
IHD	84.6%	79.0%	91.9%	£11,547	£8,521	£14,945	£9,767	£7,199	£12,944
Stroke	67.7%	57.2%	80.9%	£10,044	£6,382	£14,900	£6,805	£4,383	£10,563
HF	78.6%	70.6%	88.7%	£4,058	£2,394	£6,033	£3,191	£1,847	£4,915
Amputation	91.5%	83.9%	100%	£10,428	£7,541	£14,789	£9,546	£6,994	£13,634
Blindness 1 eye	29.9%	19.6%	49.1%	£4,533	£1,599	£8,665	£1,355	£488	£3,059
Fatal MI	96.9%	94.0%	99.4%	£1,570	£801	£2,777	£1,521	£769	£2,735
Fatal IHD	95.4%	91.1%	100%	£3,949	£1,820	£6,062	£3,766	£1,739	£5,877
Fatal Stroke	100%	100%	100%	£3,954	£2,122	£6,366	£3,954	£2,122	£6,366
History									
MI	27.0%	25.0%	32.4%	£4,280	£3,560	£5,282	£1,154	£964	£1,557
IHD	24.5%	22.8%	30.1%	£4,953	£4,390	£6,073	£1,215	£1,083	£1,672
Stroke	21.3%	19.1%	28.6%	£5,282	£3,832	£6,575	£1,125	£829	£1,677
HF	29.9%	25.0%	38.1%	£4,929	£3,335	£6,630	£1,473	£904	£2,214
Amputation	39.6%	28.5%	52.1%	£4,527	£2,828	£6,042	£1,792	£908	£2,736
Blindness 1 eye	19.6%	16.5%	25.7%	£2,311	£1,825	£3,096	£453	£336	£683

Values obtained through 1000 non-parametric bootstraps.

Table 3.20: Estimated probability of incurring some hospital in-patient costs (1), estimated annual hospital in-patient costs conditional on cost being incurred (2), and expected mean cost (3), for a representative individual (Male, age 75) during year of complication and in subsequent years

	(1) hosp			(2) cost			(3) hosp*cost		
	Coef.	**95% CI L**	**95% CI U**	**Coef.**	**95% CI L**	**95% CI U**	**Coef.**	**95% CI L**	**95% CI U**
no event	26.1%	24.4%	27.7%	3,420.96	3,131.39	3,715.99	893.74	794.26	987.58
Event during year indicator									
mi	96.4%	93.7%	99.2%	6,872.58	4,975.69	9,089.62	6,626.54	4,789.28	8,796.93
ihd	90.7%	86.0%	95.3%	11,592.32	8,724.49	15,523.22	10,510.24	7,871.12	13,998.99
stroke	78.9%	69.5%	88.3%	9,300.13	6,067.94	13,424.95	7,336.79	4,741.09	11,065.38
hf	86.6%	80.1%	93.2%	4,086.10	2,682.09	5,786.71	3,540.27	2,350.12	5,026.94
amp	94.4%	89.6%	99.1%	9,878.15	7,263.35	12,839.98	9,322.67	6,807.32	12,355.65
blind	45.5%	30.8%	60.2%	5,587.58	2,683.13	9,415.12	2,543.57	1,073.45	4,473.74
fatal mi	98.0%	96.3%	99.6%	2,659.36	1,518.42	3,819.43	2,605.64	1,489.49	3,737.42
fatal ihd	97.0%	94.5%	99.5%	4,020.66	2,554.69	5,602.93	3,899.51	2,501.56	5,468.61
fatal stroke	100%	-	-	3,795.67	2,404.48	5,287.58	3,795.67	2,404.48	5,287.58
History									
mi	39.6%	35.4%	43.9%	4,570.16	3,816.40	5,392.13	1,811.98	1,460.48	2,238.17
ihd	37.4%	33.4%	41.4%	5,313.16	4,523.52	6,221.49	1,986.49	1,621.07	2,439.70
stroke	34.4%	28.9%	39.9%	5,207.40	4,033.53	6,611.51	1,790.24	1,266.33	2,367.34
hf	42.8%	35.6%	50.0%	4,763.25	3,459.27	6,290.92	2,037.49	1,348.91	2,831.89
amp	52.3%	39.6%	65.0%	4,591.97	3,315.25	5,949.49	2,401.05	1,487.82	3,557.23
blind	30.1%	24.7%	35.6%	3,018.95	2,481.75	3,645.05	909.74	679.18	1,195.06

Values obtained through 1000 non-parametric bootstraps

Table 3.19 and Table 3.20 report predictions from the logistic regression and generalized estimating equation with gamma distribution, for an individual with characteristics set to reflect some specific values for the UKPDS study population at follow up (i.e., Table 3.19 shows results for a male aged 60 years and Table 3.20 for a male aged 75). The estimates reported are obtained through non-parametric bootstrap with 1,000 replications. This was done by randomly re-sampling individuals with replacement and re-estimating the equations and predictions. Results were then rank-ordered and the 25^{th} and 975^{th} estimates of predictions were taken as the confidence limits.

From the first column of Table 3.19 and Table 3.20 patients without complications have a probability of incurring any hospital in-patient costs in any single year of 0.16 if they are 60 years old and 0.26 if they are 75 years old (much higher than the comparable 0.06 found in Clarke et al (2003) using the first 20 years of clinical records (1977-1997) for a representative male patient aged 58). The immediate impact of complications such as MIs is to increase the marginal probability of incurring hospital in-patient costs. Thus approximately all fatal and non-fatal contemporaneous events incur a hospital in-patient costs in the year in which they occur. Patients with a history of complications also appear to have a higher probability of hospitalizations compared to those who have not experienced prior complications.

The second column of Table 3.19 and Table 3.20 show the estimated mean costs associated with each complication conditional on hospitalization. Ischemic heart diseases and amputations are associated with the greatest annual expenditures. The cost of a 60 year old male hospitalized in the year in which he had a non-fatal IHD averaged £11,547 (£8,521; £14,945), while a patient who experienced an amputation on

average incurred hospital in-patient costs of £10,428 (£7,541; £14,789) in that year. For all complications except fatal MI, these hospital in-patient costs are substantially higher than those incurred on average by the proportion of patients with no complication who nevertheless were admitted to hospitals.

The last column of Table 3.19 and Table 3.20 shows the product of the previous two columns, i.e., the expected unconditional hospital expenditure associated with each complication. Note that the 95% confidence intervals around these estimates are not symmetrical, due to the effects of retransformation.

Results indicate the substantial impact of diabetes- related complications on hospital costs, not only in the year in which the event occurs, but in permanently raising the average level of hospital costs in all subsequent years. Compared with previously reported estimates (using UKPDS data from 1977 to 1997) on the same patients at first complication, this study shows, ten years later an almost threefold increase in costs in the absence of complications alone (from £459 to £157) per annum for an hypothetical patient with the same characteristics as those in Clarke et al. (2003).

Non-inpatient costs

I now turn to the results from the six questionnaires administered between 1997 and 2007 to provide information on non-inpatient costs. These results are based on 3,589 responders and 10,920 questionnaires.

Results for the two-part model for non-inpatient costs are shown in the two tables of Appendix A3.9 (one devoted to showing coefficients and the other values in pounds for representative patients). The demeaned current age, gender, and an indicator equal to 1 if the questionnaire is any other than the first one were used as

controls. The inclusion of an indicator for whether the patient was first recruited in an English centre or not was also tested (to make results comparable with the hospitalizations records collected for English patients only), but no differences were found across nations in terms of resource utilization.

Because as illustrated in Section 3.3, only 7.2% of questionnaires report no contact with healthcare providers and so incur no costs, there is no compelling reason to model non-users separately and thus resort to a 2PM, although this proportion varies between the first and subsequent rounds. Unlike hospital costs, virtually all patients, regardless of their complications status used non-inpatient resources. Therefore the distinction between users and non-users is made redundant by the fact that after the first questionnaire almost all patients (except four) used non-inpatient resources. Our estimation equation for non-inpatient costs is:

$$E(C_{it}) = \beta^{j} X_{it}^{j},\ C_{it} \sim \text{gamma}$$

While one cannot reject the hypothesis that the coefficients are equal to each other in the first part equation ($chi^2(7) = 7.10$, $Prob > chi^2 = 0.4186$) – that is, the probability of incurring a cost is not greatly different if the complication has occurred in the period of the recall or in previous periods – in part 2 the coefficients of short term and long term are statistically different ($chi2(7) = 17.76$, $Prob > chi2 = 0.0131$). Therefore, unlike previous published results it is now possible thanks to five more rounds of questionnaires to differentiate between acute and state costs in the non-inpatient setting. The marginal impact of complications on short-term and long terms non-inpatient costs are statistically significant for all the clinical endpoints.

Table 3.21 shows the results of a GEE identity link model with gamma distribution: $(C_{it}|C_{it} > 0, X_{it}) = \beta^{j} X_{it}^{j},\ var[y|x] = (E[y|x])^2$. Standard errors are clustered at the

patient level. Table 3.22 translates those marginal effects into predicted costs for representative patients who answered at least one other questionnaire besides the first one.

I find that the biggest non-inpatient costs for any representative patient corresponds to that of an amputation. The cost is similar to that of a patient with a history of amputation. The second biggest marginal cost is due to stroke and MI.

As Tables 3.21 indicates, the age and sex of a patient are also significant explanatory variables when considering the cost of complications.

Table 3.21: Costs of non-in-patient care (GEE model)

	Coeff.	SE
Constant	534.51**	19.48
Age-65	4.61**	0.98
Male =1	-162.73**	21.51
Q>1	217.72**	18.41
Event during year indicator		
MI	465.88**	165.73
IHD	250.80*	121.64
Stroke	554.91*	252.37
HF	409.74	257.55
Amputation	1,337.21*	526.71
Blindness 1 eye	553.97	292.94
Cataract ext.	159.38*	75.06
History		
MI	132.61**	45.19
IHD	97.68*	41.29
Stroke	249.75**	81.73
HF	449.20**	121.06
Amputation	1,104.74**	276.61
Blindness 1 eye	230.83**	77.23
Cataract ext.	146.84**	41.95
Observations	10920	
Number of pts.	3589	

** p<0.01, * p<0.05

Q= questionnaire, Q>1, questionnaire other than the first one

Table 3.22: Expected costs for a representative patient (GEE model)

	60 year old Male, Q>1			75 year old Male, Q>1		
	Coef.	**95% CI L**	**95% CI U**	**Coef.**	**95% CI L**	**95% CI U**
no event	541.95	501.39	585.79	636.54	592.61	683.71
Event during year indicator						
MI	958.81	806.51	1,139.87	1,126.15	953.10	1,330.63
ihd	772.36	650.05	917.67	907.16	764.06	1,077.06
stroke	926.20	698.16	1,228.73	1,087.86	821.79	1,440.08
hf	847.78	659.01	1,090.63	995.75	778.67	1,273.34
amp	1,408.15	1,011.67	1,960.01	1,653.91	1,187.37	2,303.77
blind 1 eye	943.34	519.71	1,712.30	1,107.99	611.04	2,009.09
cataract ext.	645.89	576.08	724.16	758.61	683.70	841.74
History						
MI	655.74	581.21	739.83	770.19	688.82	861.18
ihd	639.96	564.24	725.85	751.66	665.02	849.59
stroke	749.35	613.88	914.71	880.13	728.33	1,063.58
hf	861.32	712.04	1,041.91	1,011.65	839.96	1,218.44
amp	1,376.78	1,040.07	1,822.50	1,617.08	1,222.62	2,138.80
blind 1 eye	722.99	558.92	935.23	849.18	659.48	1,093.43
cataract ext.	633.91	542.95	740.10	744.54	640.81	865.07

3.7 Conclusion and Discussion

The results presented in this chapter provide estimates of the immediate and long-term healthcare costs associated with seven major diabetes-related complications. The results cover inpatient and day-case costs, but also costs incurred in non-inpatient setting such as primary care.

Relatively few studies have analyzed the impact of complications on health care costs using patient-level data. In the UK most studies are based on population prevalence using a bottom-up approach (Dixon, Currie et al. 2000). Clarke et al. (2003) provide one UK exception: their estimates of resource use are based on participating patients in the UKPDS up to 1996/7, and their results have been widely used. The results reported here use more recent data from the UKPDS study (1998-2007), use Hospital Episode Statistics data and an HRG costing approach, and use data from six waves of patient-completed questionnaires rather than the single questionnaire used by Clarke et al (2003).

The results reported in this chapter indicate the substantial impact of many diabetes-related complications on hospital costs, not only in the year in which the event occurs, but in permanently raising the average level of hospital and non-hospital costs in all subsequent years. This effect could not easily be observed in the absence of patient-specific data, and may arise because patients carrying a history of diabetes-related complications become more expensive and difficult to treat for other health care problems when these occur.

While non-inpatient resource utilization represents roughly 30% of total expenditure, the proportion of costs attributable to non-inpatient care has steadily risen over time. The average annual costs for those with complications almost doubled from

1997 to 2007.The biggest increase in costs occurred between 1997 and 2003. This period corresponds with the publication of the UKPDS results that demonstrated the benefits of intensive therapy and the subsequent changes in treatment guidance and protocols. Table 3.23 illustrates the relation between inpatient and non-inpatient costs together with the short-term and long-term cost of a non-fatal MI, amputation and blindness (the most prevalent, most costly and least costly complications, respectively) for males and females at different ages between 50 and 80.

The panel structure of the dataset used allowed me to test cost variations emanating from heterogeneity between patients, and from patient-specific increases in the underlying utilization resulting from diabetes-related complications. I found that the sample that contributes to the identification of FE is systematically different from the sample as a whole, being predominately more likely to have complications and other causes of hospitalization. In the case of inpatient costs evidence of FE is not encountered while in the case of non-inpatient utilization, the hypothesis of a common intercept is rejected in the subset of patients that contribute to the FE identification. This is perhaps intuitive insofar as hospitalizations are dominated by non-elective cases, whereas in the case of GP visits, there is a stronger component of patient-specific choice and preferences and so different underlying levels of utilization are present across patients, irrespective of the likelihood of events. In a comparable sample, within patient differences lead to slightly lower marginal effects of diabetes complications on costs, as patients with diabetes-related events are on a higher cost path at least two years before a complication occurs. It is also important to point out that models that exploit within-patient variation tend to have wider confidence intervals and thus less precision than PA models. I argue herein that the use and interpretation of these different models ought

to be purpose-specific. For budgetary predictions one may want to trade off accuracy in favour of precision. Precise, non-causal estimates may be useful also in the targeting of cost-mitigation programs to those who appear likely to have high costs in the future. From the point of view of accepting new drugs and technologies there is here some evidence to suggest that we ought to view diabetes complications as deterministic. Once we allow for a mechanism to exist that would reduce the likelihood of events, one should be interested in causal effects and thus in isolating all other confounders of costs, in particular the non-measurable or unobservable patient-specific characteristics that lead to different rates and types of medical expenditure.

The use of long term follow-up data has allowed me to quantify the short and long term impact of different complications on costs and explore heterogeneity in the sample. It has also shown that every 10 years the cost of diabetes approximately doubles and that over time non-inpatient resource utilization has increased faster than inpatient care, possibly due to changes in the treatment protocols.

In future research, the work in this chapter could be further extended in a number of ways. The importance of multiple intercepts should be tested in different populations and in particular, the dynamic nature of expenditures and their power to predict future inpatient and non-inpatient utilization.

Conditional on data availability, there are a number of important diabetes complications that could be analysed beyond those included here such as renal disease/failure, foot ulcer, and other eye problems treated by retinal photocoagulation.

Finally, it is necessary to caution the reader on the external validity and generalizability of the findings. The UKPDS post-trial monitoring study was comprised of patients previously recruited into a large randomized trial and possibly receiving a

higher level of care than the average person with type 2 diabetes by virtue of participating in such a study. In addition, the latest data available on these patients is for the year 2007, and further changes in disease management may have occurred since then. International variations in disease management and health care costs are substantial, and these costs may not reflect the cost impact of diabetes-related complications in other health care settings. Despite these important caveats, our data have the virtues of good event adjudication, detailed clinical history, a reasonable sample size across a large number of centres, low loss to follow-up, and comprehensive and repeated measures of resource use, and should be of value to other economists and health service researchers, particularly those involved in estimating the cost-effectiveness of interventions that prevent complications.

.

Table 3.23: The Cost of a Non-Fatal MI, Amputation and Blindness in one eye for different patient characteristics

	Short Term					Long Term				Ratio non-inpatient/inpatient			
	inpatient		non-inpatient			inpatient		non-inpatient		Short Term		Long Term	
MI	Male	Female	Male	Female	MI	Male	Female	Male	Female	Male	Female	Male	Female
50	£5,858	£6,114	£986	£1,149	50	£766	£888	£653	£816	14%	16%	46%	48%
60	£6,379	£6,627	£1,032	£1,195	60	£1,154	£1,319	£699	£862	14%	15%	38%	40%
70	£6,863	£7,103	£1,078	£1,241	70	£1,667	£1,878	£745	£908	14%	15%	31%	33%
80	£7,317	£7,551	£1,124	£1,287	80	£2,300	£2,554	£791	£954	13%	15%	26%	27%
Amputation													
50	£8,821	£9,137	£1,858	£2,020	50	£1,256	£1,432	£1,625	£1,788	17%	18%	56%	56%
60	£9,546	£9,838	£1,904	£2,066	60	£1,792	£2,013	£2,324	£1,834	17%	17%	56%	48%
70	£10,181	£10,453	£1,950	£2,112	70	£2,439	£2,698	£2,369	£1,880	16%	17%	49%	41%
80	£10,744	£11,001	£1,996	£2,159	80	£3,163	£3,447	£2,415	£1,926	16%	16%	43%	36%
Blindness in one eye													
50	£915	£1,054	£1,074	£1,237	50	£269	£331	£751	£914	54%	54%	74%	73%
60	£1,355	£1,540	£1,120	£1,283	60	£453	£544	£797	£960	45%	45%	64%	64%
70	£1,923	£2,154	£1,167	£1,329	70	£723	£851	£843	£1,006	38%	38%	54%	54%
80	£2,606	£2,875	£1,213	£1,375	80	£1,095	£1,263	£889	£1,052	32%	32%	45%	45%

Appendix

A3.1 : Variables needed for Casemix Grouper

field name	description
PROCODET	Provider Code.
EPIKEY	Record identifier.
EPIORDER	Episode Number.
STARTAGE	StartAge.
SEX	Sex.
CLASSPAT	Patient Classification.
ADMISORC	Source Of Admission (Hospital Provider Spell).
ADMIMETH	Admission Method (Hospital Provider Spell).
DISDEST	Discharge Destination (Hospital Provider Spell).
DISMETH	Discharge Method (Hospital Provider Spell).
EPIDUR	Episode Duration.
MAINSPEF	Main Specialty Code.
TRETSPEF	Treatment Function Code.
DIAG_01 - DIAG_15	Additional diagnoses.
OPER_01 - OPER_12	Procedure Codes
CRITICALCAREDAYS	Critical Care Days.
REHABILITATIONDAYS	Rehabilitation Days.
SPCDAYS	Specialist Palliative Care Days.

A3.2: UKPDS 1996/7 resource use form

Box 3 COPY

UKPDS health service use questionnaire

We are trying to find out how often our patients have seen other doctors and nurses (apart from those in this clinic) during the past four months.

Would you please fill in this simple questionnaire and return it to a member of the clinic staff? Please answer every question, even if the answer is "none".

Please fill in the boxes as follows: "none" = [0][0]; "once" = [0][1]; "twice" = [0][2], and so on.

Please ask a member of staff if you have any questions about this form.

Card no: [9][9][Q] Patient no: [0][0][0][0][1][M] Visit no: [1][8][0] Date: [0][1][0][2][9][6]

During the past four months, how many times have you:

Been seen at home by:	Visited:	Telephoned for advice:	
[0][0]	[0][1]	[0][0]	your GP
[0][0]	[0][1]	[0][0]	a nurse
[0][0]	[0][0]	[0][0]	a health visitor
[0][0]	[0][0]	[0][0]	a dietitian
[0][0]	[0][0]	[0][0]	a chiropodist
	[0][0]	[0][0]	an optician
	[0][1]	[0][0]	a hospital eye clinic
	[0][0]	[0][0]	any other hospital clinic

Do you have any other comments that you would like to make?

Like to sleep a lot,

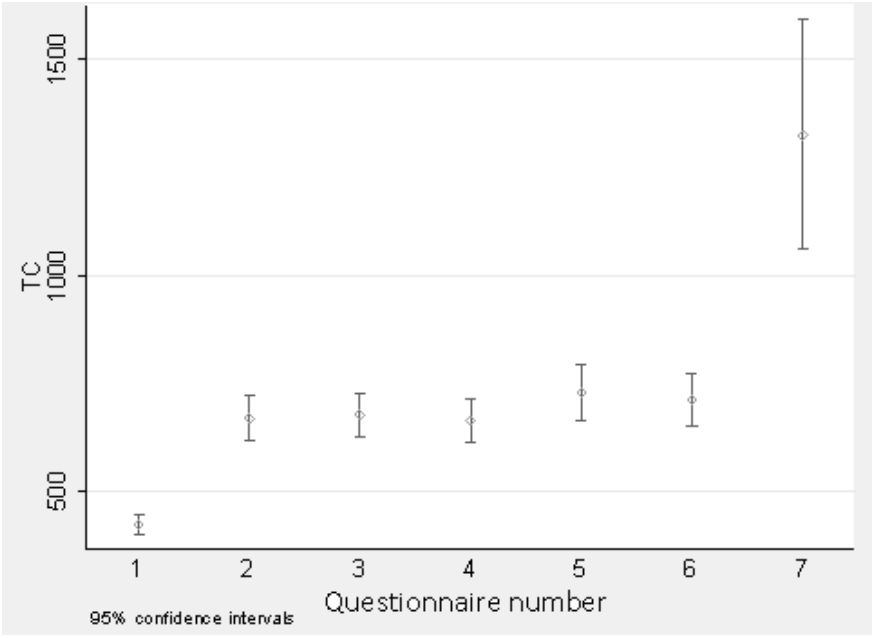
A3.3: Response Rate questionnaires 2 to 7

	2002/3			2003/4			2004/5			2005/6			2006/7			Oct-07					
corresponding city	I	R	RATE	I	R	%	I	R	%	I	R	%	I	R	%	I	R	RATE closeout	**TOTAL ISSUED**	**TOTAL RETURNED**	**RATE**
Aberdeen	94	70	74%	97	77	79%	91	67	74%	86	67	78%	67	48	72%	73	52	71%	508	381	75%
Belfast City	62	37	60%	60	42	70%	57	36	63%	54	33	61%	43	24	56%	51	31	61%	327	203	62%
Belfast Victoria	126	79	63%	123	89	72%	120	75	63%	117	71	61%	87	56	64%	106	61	58%	679	431	63%
Birmingham	122	90	74%	113	85	75%	109	74	68%	103	71	69%	85	54	64%	98	65	66%	630	439	70%
Carshalton	103	84	82%	101	82	81%	98	81	83%	96	79	82%	69	59	86%	85	74	87%	552	459	83%
Derby	79	48	61%	78	52	67%	78	50	64%	75	39	52%	51	23	45%	69	38	55%	430	250	58%
Dundee	115	97	84%	114	96	84%	111	93	84%	107	89	83%	83	67	81%	99	86	87%	629	528	84%
Hammersmith	84	51	61%	82	37	45%	78	43	55%	74	38	51%	55	27	49%	67	37	55%	440	233	53%
Ipswich	113	93	82%	109	89	82%	99	86	87%	98	81	83%	76	61	80%	86	72	84%	581	482	83%
Leicester	51	42	82%	49	41	84%	48	37	77%	44	32	73%	31	24	77%	43	29	67%	266	205	77%
London	108	66	61%	102	67	66%	98	67	68%	97	62	64%	72	53	74%	89	60	67%	566	375	66%
Manchester	41	34	83%	45	37	82%	45	36	80%	44	34	77%	26	22	85%	39	31	79%	240	194	81%
Northampton	121	105	87%	115	96	83%	111	97	87%	106	80	75%	79	63	80%	95	83	87%	627	524	84%
Norwich	124	100	81%	118	100	85%	112	93	83%	105	88	84%	80	62	78%	99	74	75%	638	517	81%
Oxford	180	144	80%	174	149	86%	166	141	85%	164	132	80%	128	95	74%	151	121	80%	963	782	81%
Peterborough	41	38	93%	40	37	93%	39	35	90%	37	33	89%	28	25	89%	33	31	94%	218	199	91%
Salford	103	60	58%	103	62	60%	102	57	56%	99	52	53%	83	51	61%	99	56	57%	589	338	57%
Scarborough	96	70	73%	96	77	80%	94	67	71%	94	72	77%	80	47	59%	74	52	70%	534	385	72%
St. George's	244	160	66%	234	149	64%	225	140	62%	219	127	58%	164	92	56%	199	111	56%	1285	779	61%
Stevenage	71	63	89%	69	61	88%	67	61	91%	65	58	89%	50	46	92%	61	49	80%	383	338	88%
Stoke-on-Trent	105	80	76%	100	86	86%	97	81	84%	92	73	79%	72	59	82%	81	63	78%	547	442	81%
Torbay	63	39	62%	61	46	75%	60	44	73%	60	43	72%	50	37	74%	56	40	71%	350	249	71%
Whittington	92	75	82%	88	74	84%	85	72	85%	83	66	80%	66	50	76%	77	64	83%	491	401	82%
TOTAL	**2338**	**1725**	**74%**	**2271**	**1731**	**76%**	**2190**	**1633**	**75%**	**2119**	**1520**	**72%**	**1625**	**1145**	**70%**	**1930**	**1380**	**72%**	**12473**	**9134**	**73%**

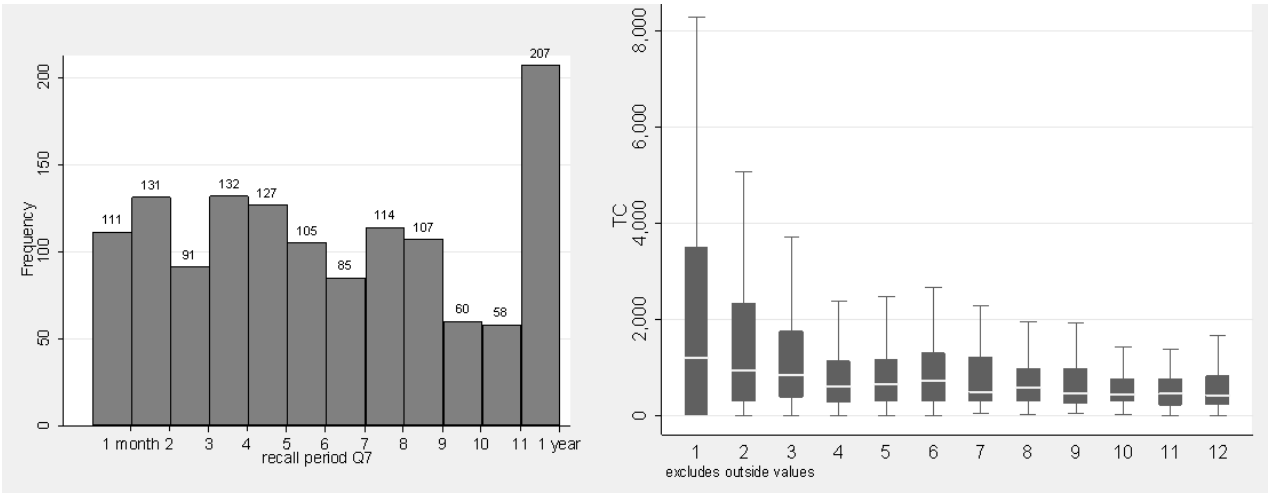
I=issue, R=returned

A3.4: Analysing questionnaire seven

Comparing average total cost and CI by questionnaire number



Frequency of recall period and median total cost per recall period for questionnaire 7



A3.5: Unit costs of non-inpatient health services

Resource	Seen at home	Surgery visit	Telephone advice	Source
General Practitioner (exc. Drugs)	£4.70 per minute or **£110** per home visit lasting 23.4 minutes (includes travel time of 12 min)	**£40** per consultation lasting 11.7 minutes, excluding direct care staff costs	**£24** for 7.1 minutes	Lesley Curtis PSSRU 2012. Table 10.8b, page 183
Nurse (clinical nurse specialist)	**£24** per home visit including qualifications and travel	**£14** per consultation	6 min. £58 per hour= **£5.8**	Lesley Curtis PSSRU 2012. Table 10.6, page 180.
Health visitor	**£30** per home visit (Assume 25 minutes)	£59 per hour of clinic contact. Assume 15 min. **£15**	£59 per hour. Assume 6 min. **£6**	Lesley Curtis PSSRU 2012. Table 10.3, page 177
Dietician	Assume health visitor costs: **£30** per home visit	Assume health visitor costs: **£15** per surgery visit	Assume health visitor costs: **£6** per telephone advice	Lesley Curtis PSSRU 2012. Table 10.3, page 177
Chiropodist	Assume health visitor costs: **£30** per home visit	Assume health visitor costs: **£15** per surgery visit	Assume health visitor costs: **£6** per telephone advice	Lesley Curtis PSSRU 2012. Table 10.3, page 177
Optician	NHS Sight Test Fees for 2012 **£36.10**	NHS Sight-test fees **£20.69** agreed for 2012	**£5.17** (assumed to be 25% of surgery consultation)	Department of Health & Federation of Opticians http://www.fodo.com/resource-categories/nhs-sight-test-fees
Hospital eye clinic		**£80**		National Schedule of Reference Costs Year: 2011-12 - All NHS trusts and NHS foundation trusts - HRG Data. Average cost per out-patient ophthalmology consultation, consultant-led, follow-up, 2011-12
Any other hospital clinic		**£115**		National Schedule of Reference Costs Year: 2011-12 - All NHS trusts and NHS foundation trusts - HRG Data. Average cost per out-patient consultation, consultant-led, follow-up, all except ophthalmology, 2011-12

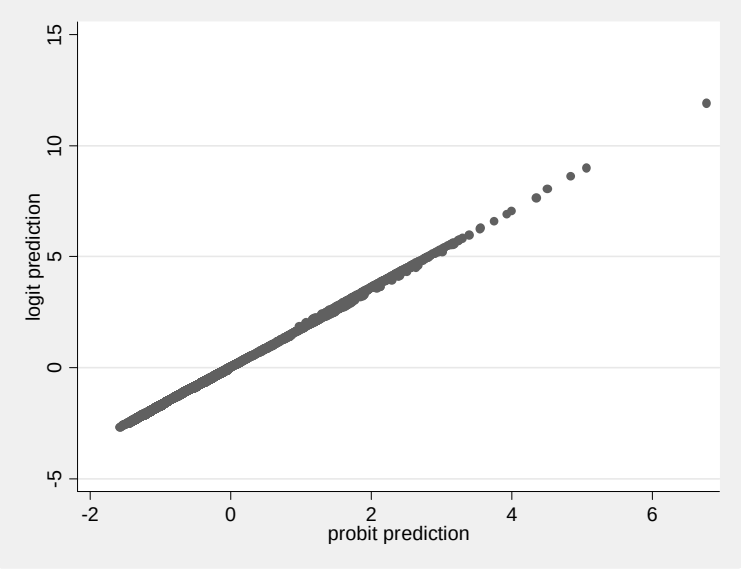
A3.6: Comparing alternative models for the first stage in the 2PM

hospitalization	(xtprobit) P(z-score)	(xtlogit) log(p)/log(1-p)	(RE) LPM
Constant	-0.741**	-1.253**	0.273**
	(0.027)	(0.046)	(0.007)
Age-67	0.025**	0.043**	0.007**
	(0.002)	(0.003)	(0.000)
Male =1	-0.070*	-0.120*	-0.017
	(0.034)	(0.058)	(0.009)
Event during year indicator			
mi	2.539**	4.495**	0.605**
	(0.222)	(0.451)	(0.038)
ihd	1.960**	3.366**	0.547**
	(0.154)	(0.282)	(0.035)
stroke	1.392**	2.394**	0.397**
	(0.162)	(0.286)	(0.041)
hf	1.734**	2.960**	0.476**
	(0.169)	(0.303)	(0.038)
amp	2.279**	4.063**	0.523**
	(0.278)	(0.556)	(0.047)
blind	0.494**	0.847**	0.150**
	(0.169)	(0.284)	(0.048)
fatal mi	2.877**	5.095**	0.602**
	(0.264)	(0.539)	(0.034)
fatal ihd	2.651**	4.682**	0.566**
	(0.270)	(0.544)	(0.038)
History			
mi	0.399**	0.676**	0.113**
	(0.058)	(0.099)	(0.016)
ihd	0.334**	0.563**	0.092**
	(0.054)	(0.092)	(0.015)
stroke	0.252**	0.434**	0.076**
	(0.072)	(0.123)	(0.021)
hf	0.487**	0.826**	0.141**
	(0.084)	(0.143)	(0.024)
amp	0.748**	1.276**	0.193**
	(0.128)	(0.218)	(0.035)
blind	0.151*	0.260*	0.047*
	(0.072)	(0.122)	(0.020)
Observations	20889		
Number of ukno	3074		
R-squared			

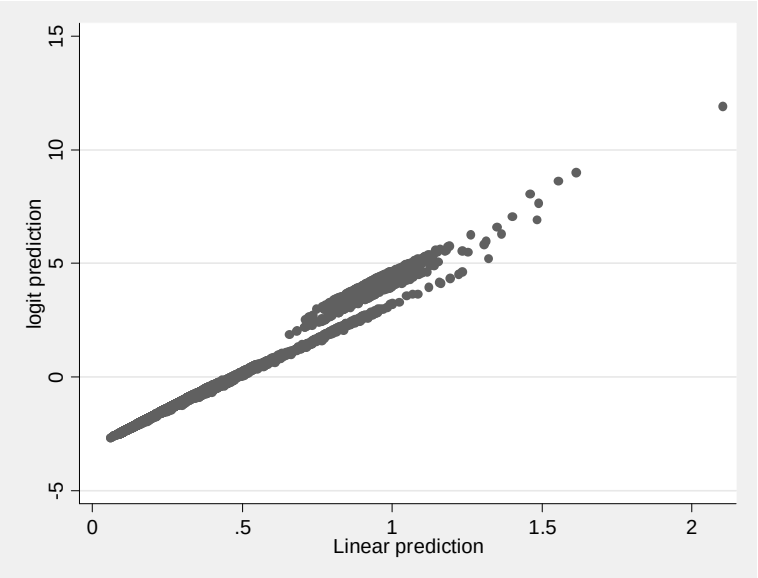
Standard errors in parentheses
** p<0.01, * p<0.05

Interpretation of the coefficients in probit regression is not as straightforward as the interpretations of coefficients in linear regression or logit regression. The increase in probability attributed to a one-unit increase in a given predictor is dependent both on the values of the other predictors and the starting value of the given predictors (A one unit change in X leads to a β change in the z-score of the likelihood of hospitalizatio).

Logistic regression is generally preferred over the probit model because the exponentiated logit coefficients can be interpreted as odds ratios -- which is not the case with probit coefficients.



Model	N	AIC	BIC
Probit	20889	22538.19	22681.23
Logit	20889	22541.55	22684.6



A3.7: Test for link function

Inpatient costs

link(log) family(gamma)

```
Variance function: V(u) = 1                          [Gaussian]
Link function    : g(u) = u                          [Identity]

                                                     AIC                =   20.9878
Log likelihood   = -67315.37176                      BIC                =  4.89e+11
```

tot_cost	OIM Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_hat	-17432.32	8840.723	-1.97	0.049	-34759.82	-104.8234
_hatsq	1333.715	511.1441	2.61	0.009	331.8908	2335.539
_cons	56957.19	38148.98	1.49	0.135	-17813.43	131727.8

link(identity) family(gamma)

```
Variance function: V(u) = 1                          [Gaussian]
Link function    : g(u) = u                          [Identity]

                                                     AIC                =  20.98538
Log likelihood   = -67307.62165                      BIC                =  4.88e+11
```

tot_cost	OIM Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_hat	.7533513	.1960361	3.84	0.000	.3691276	1.137575
_hatsq	.0000153	.0000133	1.15	0.251	-.0000108	.0000414
_cons	706.0675	574.4812	1.23	0.219	-419.8949	1832.03

Non-inpatient costs

link(log) family(gamma)

```
Variance function: V(u) = 1                          [Gaussian]
Link function    : g(u) = u                          [Identity]

                                                     AIC                =  16.73605
Log likelihood   = -91375.81679                      BIC                =  1.19e+10
```

TC	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
_hat	-3142.663	532.0054	-5.91	0.000	-4185.375	-2099.952
_hatsq	296.2111	40.24906	7.36	0.000	217.3244	375.0978
_cons	8586.681	1756.032	4.89	0.000	5144.922	12028.44

link(identity) family(gamma)

```
Variance function: V(u) = 1                          [Gaussian]
Link function    : g(u) = u                          [Identity]

                                                     AIC                =  16.73441
Log likelihood   = -91366.85322                      BIC                =  1.18e+10
```

TC	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
_hat	1.125342	.0844237	13.33	0.000	.9598748	1.29081
_hatsq	-.0000557	.000034	-1.64	0.102	-.0001224	.000011
_cons	-53.90664	42.09182	-1.28	0.200	-136.4051	28.5918

If the model is specified correctly then the prediction squared would have no explanatory power (McCullagh and Nelder 1989).

A3.8: Diagnostics – Distribution Family (Park Test)

```
glm tot_cost age_cent  male  $events_65 $history_65 if hosp==1 , family(gamma) link(log)
predict predcosts, mu
gen ressq=(tot_cost-predcosts)^2
gen ln_predcosts=ln(predcosts)
glm ressq ln_predcosts if hosp==1, family(gamma) link(log) robust nolog
```

The implementation of this test requires a regression of the log of the predicted value of costs *(tot_cost)* on the square of the residuals and a constant and test for the value (=0,1,2,3) of the coefficient on log of the predicted value of hospital costs.

```
. glm ressq ln_predcosts if hosp==1, family(gamma) link(log) robust nolog

Generalized linear models                         No. of obs      =      6415
Optimization     : ML                             Residual df     =      6413
                                                  Scale parameter =  34.57692
Deviance         =  25530.82196                   (1/df) Deviance =  3.981104
Pearson          =  221741.8046                   (1/df) Pearson  =  34.57692

Variance function: V(u) = u^2                     [Gamma]
Link function    : g(u) = ln(u)                   [Log]

                                                  AIC             =  37.87656
Log pseudolikelihood = -121487.0537               BIC             = -30688.06

------------------------------------------------------------------------------
             |               Robust
       ressq |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ln_predcosts |   1.742727   .1775876     9.81   0.000     1.394662    2.090793
       _cons |   3.441994   1.497011     2.30   0.021     .5079066    6.376081
------------------------------------------------------------------------------

.
. test ln_predcosts==0  // gaussian

 ( 1)  [ressq]ln_predcosts = 0

           chi2(  1) =   96.30
         Prob > chi2 =    0.0000

. test ln_predcosts==1 // poisson

 ( 1)  [ressq]ln_predcosts = 1

           chi2(  1) =   17.49
         Prob > chi2 =    0.0000

. test ln_predcosts==2 // gamma

 ( 1)  [ressq]ln_predcosts = 2

           chi2(  1) =    2.10
         Prob > chi2 =    0.1474

. test ln_predcosts==3  // inverse gamma

 ( 1)  [ressq]ln_predcosts = 3

           chi2(  1) =   50.12
         Prob > chi2 =    0.0000
```

A3.9: Two-Part model for non-inpatient costs

Table 1: Regression equations to derive annual probability of incurring non-inpatient costs, and costs of conditional outpatient care

	Part 1: Equation to derive annual probability of incurring non-inpatient costs		**Part 2: Equation to derive annual cost of non-inpatient care, conditional on costs being incurred**	
	RE Logit		**GLM Gamma distribution log identity**	
Variable	*Coefficient*	*SE*	*Coefficient*	*SE*
Constant	2.192**	(0.163)	618.463**	(24.888)
Age-65	0.023**	(0.005)	3.493**	(1.283)
Male (=1)	-0.636**	(0.104)	-153.232**	(28.203)
Questionnaire number >1	3.075**	(0.157)	97.489**	(26.931)
Event during year indicator				
Non-fatal MI	1.788	(1.070)	388.968**	(91.535)
Non-fatal IHD	20.368	(11,072.119)	289.478**	(70.560)
Non-fatal stroke	-0.020	(0.845)	563.691**	(198.201)
Heart failure	20.016	(16,448.917)	420.003**	(143.414)
Amputation	-1.615	(0.865)	2,323.388**	(731.905)
Blindness in one eye	0.291	(1.128)	1,302.265*	(614.210)
Cataract extraction	1.158	(0.647)	149.671**	(37.825)
History of event†				
Non-fatal MI	0.494*	(0.232)		
Non-fatal IHD	0.303	(0.221)	131.633**	(46.262)
Non-fatal stroke	0.060	(0.339)	122.824**	(45.739)
Heart failure	0.949	(0.563)	239.109**	(88.735)
Amputation	0.515	(0.693)	421.668**	(97.388)
Blindness in one eye	0.480	(0.337)	1,097.819**	(252.624)
Cataract extraction	-0.019	(0.227)	218.291*	(100.474)
Observations	10920		10238	
Number of patients	3589		3330	

†Refers to events that occurred in previous patient years. ** p<0.01, * p<0.05

Table 2: Estimated probability of incurring non-inpatient costs, estimated annual non-inpatient costs conditional on cost being incurred, for a representative individual, during year of complication and in subsequent years

	Probability of incurring some non-inpatient costs		Estimated annual non-inpatient costs (GEE Gamma Model with Link Function) conditional on cost being incurred		Expected non-inpatient cost of complication	
	Mean	95% CI	Mean (£)	95% CI	Mean (£)	95% CI
No Complication	0.991	(0.988, 0.995)	546	(510, 584)	541	(419, 595)
Event during year indicator						
Non-fatal MI	0.999	(0.995, 1)	934	(789, 1106)	933	(767, 1106)
Non-fatal IHD	1		744	(627, 884)	744	(592, 860)
Non-fatal stroke	0.991	(0.976, 1)	934	(730, 1195)	926	(792, 1203)
Heart failure	1		792	(614, 1022)	792	(631, 969)
Amputation	0.958	(0.889, 1)	1,378	(981, 1935)	1,320	(1001, 1535)
Blindness in one eye	0.994	(0.979, 1)	968	(544, 1722)	962	(735, 1476)
Cataract extraction	0.997	(0.994, 1)	627	(560, 702)	625	(545, 723)
History of event						
Non-fatal MI	0.995	(0.992, 0.998)	657	(589, 733)	654	(579, 765)
Non-fatal IHD	0.994	(0.990, 0.997)	644	(571, 726)	640	(563, 722)
Non-fatal stroke	0.992	(0.986, 0.998)	754	(628, 905)	748	(599, 907)
Heart failure	0.997	(0.993, 1)	866	(724, 1037)	863	(703, 1008)
Amputation	0.995	(0.988, 1)	1,352	(1009, 1811)	1,345	(876, 1576)
Blindness in one eye	0.995	(0.990, 0.999)	721	(561, 928)	718	(545, 830)
Cataract extraction	0.991	(0.986, 0.996)	653	(560, 762)	647	(575, 715)

A3.10: Testing for FE

Inpatient Costs

To ascertain whether or not a fixed effects model is appropriate, I test whether the time-invariant error term is correlated with the regressors using the methodology formulated by Hausman – the null hypothesis is that they are not and this hypothesis is rejected (χ^2=111.60, Prob> χ^2<0.002). Because FE cuts down on the sample for identification, I also show that the outcome of multiple intercepts is driven by the sub-sample that helps towards the identification of within differences.

Before showing what drives the difference in results, Table A3.10.1 reports coefficients of two specifications of the outcomes equations from a two part model, where costs are estimated using within-patient differencing. The first specification of costs conditional on hospitalizations is in levels and the second is in logs to address the skewness of the data and produce a more symmetric distribution of the conditional costs.

The interpretation of the FE results in levels is that a one-unit change in the predictor results in β units change in the expected value of the response variable, *ceteris paribus*. Interpreting a log transformed variable can also be done in such a manner. However, such coefficients are routinely interpreted in terms of percent change on the original level scales ($(e^{\beta} - 1) \times 100$). The last column of Table A3.12.1 shows the multiplicative effects of the log transformed model.

It is not straightforward to ascertain whether levels or logs are best in the context of FE. For instance goodness of fit measures like R-squared cannot be used to compare models using different transformation scales for the dependent variable -- transformations change the measure of the variance. Similarly information criteria like

AIC and BIC only work with ML estimators. In choosing the appropriate model we can then use the following three guidelines: (1) test on multiplicative effects of complications on costs (page 153), (2) test on the residual (AppendixA3.8) and (3) keep in mind what is common practice in the literature. Using these considerations, there is an overwhelming support for a model in logs. However it is important to stress that any gains in precision that would be achieved by the log transformation may be lost in the re-transformation process.

Table A3.10.1 FE in levels (1) and logs (2) conditional on hospitalization

	(1) cost		(2) log of cost		%
	Coeff	SE	Coeff	SE	change
Constant	2,957.532**	(277.008)	7.148**	(0.038)	
Age-67	158.627**	(60.428)	0.018	(0.010)	1.82%
Event during year indicator					
mi	4,485.242**	(1,328.880)	0.619**	(0.201)	85.71%
ihd	9,425.548**	(2,016.248)	1.167**	(0.185)	221.23%
stroke	5,343.862*	(2,230.405)	0.403	(0.250)	49.63%
hf	-352.014	(1,301.867)	-0.144	(0.200)	-13.41%
amp	8,613.129**	(2,267.511)	1.413**	(0.217)	310.83%
blind	2,519.740	(1,655.061)	0.577*	(0.238)	78.07%
fatal mi	-159.628	(1,384.030)	-0.498**	(0.182)	-39.23%
fatal ihd	1,006.868	(1,536.569)	-0.288	(0.217)	-25.02%
fatal stroke	-1,421.778	(2,270.856)	-0.341	(0.287)	-28.89%
History					
mi	2,358.607	(1,369.114)	0.333	(0.216)	39.51%
ihd	3,224.130*	(1,351.477)	0.346	(0.188)	41.34%
stroke	2,692.405	(2,699.538)	0.232	(0.277)	26.11%
hf	-1,024.975	(1,303.956)	-0.054	(0.244)	-5.26%
amp	-854.629	(1,998.953)	0.272	(0.280)	31.26%
blind	-699.988	(1,521.695)	0.026	(0.235)	2.63%
Observations	6415		6415		
Number of pts	2389		2389		

** p<0.01, * p<0.05

In the first stage I use a FE linear probability model (LPM)[17] because it is easy to interpret and because unlike the logit FE (also referred to as conditional logit) it allows one to predict probability in levels rather than purely marginal effects. Most importantly FE LMP does not discard observations based on the outcome of the dependent variable[18]. Because FE logit estimates use only within-individual differences, it uses subjects as their own controls. The estimation drops 902 patients either because they never go to the hospital (685) or because they are hospitalized at least once in every year we observe them (217). This complicates comparisons with population average models as the set of patients included for estimation changes. Using a linear model in both the first and the second equations allows us to maintain the same estimating sample as that used in the population-averaged models. It is important to point out however that only patients with costs before and after an event contribute to the identification of the parameters in the FE. Unlike conditional logit (FE logit) the variation in the dependent variable never determines the estimating sample but it is still the variation in the explanatory variables to determine identification. Here the non-contributors are individuals who have a complication of a given type in every year of observation or people who never have a complication.

Table A3.10.2 illustrates the prediction probabilities for two representative patients after using the FE LPM.

[17] I test for the presence of FE also in the first stage equation and reject the null hypothesis. Test: Ho: difference in coefficients between a LPM with and without FE is not systematic. Chi2(16) = 250.86 Prob>chi2 =0.0000.

[18] The conditional logit requires variation in both outcome and explanatory, whereas LPM with FE requires only variation in explanatory variables.

Table A3.10.2: Estimates of the probability of incurring hospital costs under FE LPM

	Coeff	SE
Constant	0.266**	(0.006)
Male (=1)		
Age-67	0.024**	(0.001)
Event during year indicator		
mi	0.607**	(0.042)
ihd	0.494**	(0.039)
stroke	0.354**	(0.046)
hf	0.406**	(0.042)
amp	0.429**	(0.051)
blind	0.110*	(0.054)
fatal mi	0.546**	(0.039)
fatal ihd	0.520**	(0.042)
fatal stroke	0.558**	(0.053)
History		
mi	0.117**	(0.037)
ihd	-0.001	(0.031)
stroke	0.022	(0.039)
hf	0.056	(0.044)
amp	-0.022	(0.054)
blind	-0.009	(0.046)
Observations	20889	
Number of pts	3074	

Table A3.10.3 and Table A3.10.4 show respectively the conditional and unconditional cost for representative patients of different ages after the log re-transformation using Duan's smearing estimator: $E[C|x] = \hat{\phi} \times \exp(x\hat{\beta})$, where $\hat{\phi}$ is the estimated smearing factor (the mean of the exponentiated residuals). The unconditional costs in Table A3.10.4 are again obtained by multiplying the probability of hospitalization times the conditional costs. Compared with the GEE results obtained using population averages, FE results show slightly higher in hospital costs associated with complications.

Table A3.10.3: Predicted Probabilities for representative patients after FE

	Probability of hospitalization					
	60 years old Male			75 years old male		
	Coef.	95% CI U	95% CI L	Coef.	95% CI U	95% CI L
no event	9.6%	7.8%	11.5%	46.0%	43.5%	48.4%
Event during year indicator						
mi	70.3%	62.1%	78.6%	100%	98.5%	100%
ihd	59.1%	51.3%	66.8%	95.4%	87.7%	100%
stroke	45.0%	36.0%	54.1%	81.4%	72.4%	90.4%
hf	50.2%	41.8%	58.6%	86.5%	78.3%	94.8%
amp	52.5%	42.4%	62.6%	88.8%	78.7%	98.9%
blind	20.6%	10.0%	31.3%	57.0%	46.3%	67.6%
fatal mi	64.2%	56.2%	72.2%	95.6%	92.7%	100%
fatal ihd	61.7%	53.0%	70.3%	98.0%	89.5%	100%
fatal stroke	65.8%	57.1%	74.3%	99.4%	95.6%	100%
History						
mi	21.4%	14.2%	28.5%	57.7%	51.0%	64.5%
ihd	9.5%	3.3%	15.8%	45.9%	40.1%	51.6%
stroke	11.8%	4.0%	19.6%	48.1%	40.6%	55.7%
hf	15.3%	6.5%	24.0%	51.6%	43.1%	60.1%
amp	7.4%	0.4%	18.3%	43.8%	33.1%	54.4%
blind	8.7%	0.2%	17.7%	45.1%	36.4%	53.7%

Table A3.10.4: Costs conditional on hospitalization under FE for two representative patients

	conditional on hospitalization					
	60 years old Male			75 years old male		
	Coef.	95% CI U	95% CI L	Coef.	95% CI U	95% CI L
no event	4,400.92	4,104.27	4,661.62	4,400.92	4,095.74	4,636.90
Event during year indicator						
mi	5,782.81	4,482.77	10,254.37	7,568.87	4,753.76	10,119.97
ihd	10,005.23	8,294.03	16,858.89	13,095.42	8,124.57	16,556.36
stroke	4,659.28	3,333.74	9,268.43	6,098.33	3,412.63	9,667.70
hf	2,696.91	2,283.84	4,993.71	3,529.87	2,262.37	4,759.66
amp	12,793.22	10,052.84	25,171.61	16,744.50	9,188.40	22,992.32
blind	5,544.53	4,006.93	11,453.53	7,256.99	4,050.70	10,268.03
fatal mi	1,893.01	1,564.08	3,356.18	2,477.68	1,573.05	3,215.29
fatal ihd	2,334.69	1,753.23	4,293.13	3,055.78	1,858.76	4,368.75
fatal stroke	2,213.07	1,471.45	4,653.96	2,896.59	1,567.61	5,084.41
History						
mi	4,343.28	3,712.15	7,810.58	5,684.73	3,528.78	7,976.56
ihd	4,402.83	3,810.71	7,084.07	5,762.67	3,753.45	7,479.44
stroke	3,925.58	3,022.94	8,122.18	5,138.03	2,543.71	8,610.79
hf	2,951.51	2,079.89	5,787.66	3,863.10	2,317.09	5,927.24
amp	4,088.37	2,995.62	8,556.49	5,351.09	2,690.70	8,461.76
blind	3,196.00	2,581.89	6,279.74	4,183.11	2,520.39	6,900.91

Table A3.10.5: Unconditional hospitalization costs under FE for two representative patients

	unconditional costs					
	60 years old Male			75 years old male		
	Coef.	95% CI U	95% CI L	Coef.	95% CI U	95% CI L
no event	423.45	338.41	829.00	2,023.11	829.00	2,182.14
Event during year indicator						
mi	4,066.44	2,742.02	6,593.22	7,568.87	4,807.65	11,892.37
ihd	5,908.50	3,793.57	9,186.00	12,493.36	7,943.69	15,899.93
stroke	2,098.34	1,217.80	3,763.06	4,963.07	2,837.59	8,176.67
hf	1,353.47	860.40	2,209.25	3,054.55	1,899.59	4,104.21
amp	6,714.73	4,385.85	11,581.21	14,874.97	8,563.81	20,870.10
blind	1,143.15	394.18	2,227.94	4,134.02	2,166.24	6,220.96
fatal mi	1,215.86	844.23	1,987.62	2,368.66	1,588.15	3,244.97
fatal ihd	1,439.61	936.19	2,427.64	2,994.98	1,759.21	4,282.84
fatal stroke	2,213.07	768.06	2,802.68	2,896.59	1,547.53	4,679.92
History						
mi	927.86	488.58	1,801.42	3,280.74	1,927.08	4,555.59
ihd	419.70	123.09	876.02	2,643.97	1,686.39	3,501.54
stroke	462.43	13.01	1,046.53	2,472.85	1,310.65	4,028.73
hf	450.52	133.37	1,025.13	1,993.84	1,050.93	2,873.21
amp	303.71	195.90	1,409.41	2,342.54	1,070.99	4,140.36
blind	279.46	114.65	822.40	1,886.27	995.08	3,001.47

*CI are obtained through non-parametric bootstrapping.

Is the sample that changes the coefficients or is the estimation?

In order to test for the hypothesis that results might driven by the subsample of identification I estimate an OLS regression on the sample as a whole and on the inclusive subsample for identification of FE (i.e., all those patients for which we have costs before and after at least one complication), I then compare the coefficients across these two estimations. If coefficients do not change then differences between FE and PA are to be determined by the specification rather than the sample for identification. I reject the hypothesis that the coefficients are the same across the pooled sample and the FE sample (Prob>chi2=0.024).

To further test whether FE may still need to be considered in a subsample, even if unrepresentative of the whole, I carry on a Haussmann test of the subsample of

switchers, thus holding constant the sample for identification. Here one cannot reject the null hypothesis that the coefficients of FE and RE are in fact the same (Prob>chi2=0.347).

Non-inpatient Costs

The hypothesis of a single intercept is rejected also in the case of non-inpatient utilization in the sample as a whole. Once more however the identifying sample under FE is different than the overall sample.

For completeness sake, the coefficients of FE are shown in Table A3.10.6 despite depicting an unrepresentative subsample. These are to be interpreted in terms of percentage changes as the dependent variable is expressed in logs. Because we are interested in level effects rather than percentages, in Table A3.10.7, I illustrate expected costs for two representative patients after applying Duan smearing estimator.

Table A3.10.6: Costs of non-in-patient care (FE model, log(Costs))

	Coeff	SE	% change
Constant	5.780**	(0.032)	
Age-65	0.029**	(0.006)	2.9%
Q>1	0.251**	(0.058)	28.5%
Event during year indicator			
MI	0.586**	(0.098)	79.7%
ihd	0.222*	(0.091)	24.9%
stroke	0.590**	(0.137)	80.4%
hf	0.388*	(0.155)	47.4%
amp	0.773**	(0.223)	116.6%
blind 1 eye	0.225	(0.185)	25.2%
cataract ext.	0.282**	(0.055)	32.6%
History			
MI	0.118	(0.097)	12.5%
ihd	0.031	(0.093)	3.1%
stroke	0.259	(0.137)	29.6%
hf	0.391**	(0.139)	47.8%
amp	0.683**	(0.218)	98.0%
blind 1 eye	0.122	(0.160)	13.0%
cataract ext.	0.037	(0.056)	3.8%
+Observations	10238		
+Number of pts.	3330		

** p<0.01, * p<0.05, +we lose 259 patients who report zero health utilization.

Table A3.10.7: Expected costs for a representative patient (FE model, log(Costs))

FE	60 year old Male Q>1			75 year old Male Q>1		
	Coef.	**95% CI L**	**95% CI U**	**Coef.**	**95% CI L**	**95% CI U**
no event	582.32	460.87	621.15	721.26	698.37	898.13
Event during year indicator						
MI	1,024.41	783.98	1,228.57	1,580.12	1,220.28	1,943.79
ihd	711.99	542.41	827.90	1,098.22	832.70	1,250.78
stroke	1,028.61	713.47	1,387.83	1,586.59	1,147.28	2,016.33
hf	839.98	494.43	1,114.27	1,295.64	853.61	1,616.40
amp	1,234.79	740.16	1,803.70	1,904.63	1,158.63	2,766.36
blind 1 eye	714.12	445.41	929.75	1,101.51	680.62	1,466.15
cataract ext.	755.59	595.39	815.65	1,165.48	915.67	1,269.96
History						
MI	641.69	465.29	752.38	989.79	725.15	1,136.99
ihd	588.03	425.08	668.33	907.01	671.74	999.35
stroke	738.51	501.57	922.39	1,139.13	770.31	1,449.34
hf	842.70	577.45	1,052.42	1,299.84	886.65	1,623.00
amp	1,128.85	648.27	1,689.49	1,741.21	1,088.82	2,633.19
blind 1 eye	644.15	401.31	806.78	993.58	630.79	1,259.60
cataract ext.	591.26	452.62	632.96	912.00	732.89	968.97

Testing for sample equivalence:

Just like in the case of inpatient utilization, the hypothesis that the coefficients are the same across the pooled sample and the FE sample is rejected (Prob>chi2=0.0043).

In the case of non-inpatient resource utilization however, despite the sample of identification for FE not being representative of the whole, one still rejects the hypothesis of a single intercept even when holding the sample of switchers constant across estimations (Prob>chi2=0.001).

References

ADA (2003). "Economic Costs of Diabetes in the U.S. in 2002." Diabetes Care 26(3): 917-932.

ADA (2005). "Standards of Medical Care in Diabetes." Diabetes Care 28(suppl 1): s4-s36.

Bagust, A., P. K. Hopkinson, et al. (2002). "The projected health care burden of Type 2 diabetes in the UK from 2000 to 2060." Diabetic Medicine **19**: 1-5.

Ballesta, M., F. Carral, et al. (2006). "Economic cost associated with type II diabetes in Spanish patients." The European Journal of Health Economics **7**(4): 270-275.

Basu, A. and W. G. Manning (2009). "Issues for the Next Generation of Health Care Cost Analyses." Medical Care **47**(N7 Supplement 1): S109-S114

Basu, A. and P. J. Rathouz (2005). "Estimating marginal and incremental effects on health outcomes using flexible link and variance function models." Biostat **6**(1): 93-109.

Berenson, A. and E. Rich (2010). "US Approaches to Physician Payment: The Deconstruction of Primary Care." J Gen Intern Med **25**(6): 613-8.

Bevan, G. and R. Robinson (2005). "The interplay between economic and political logics: path dependency in health care in England." Journal of Health Politics, Policy and Law (30): 53 - 78.

BMA (March 2008). Payment by Results Briefing Note. H. P. a. E. R. Unit.

Bogner, H. R., S. D. Miller, et al. (2010). "Assessment of Cost and Health Resource Utilization for Elderly Patients With Heart Failure and Diabetes Mellitus." Journal of Cardiac Failure **16**(6): 454-460.

Briggs, A., R. Nixon, et al. (2005). "Parametric modelling of cost data: some simulation evidence." Health Economics **14**(4): 421 - 428.

Brown, J. B., K. L. Pedula, et al. (1999). "The Progressive Cost of Complications in Type 2 Diabetes Mellitus." Arch Intern Med **159**(16): 1873-1880.

Clarke, P., A. Gray, et al. (2003). "The impact of diabetes-related complications on healthcare costs: results from the United Kingdom Prospective Diabetes Study (UKPDS Study No. 65)." Diabetic Medicine **20**(6): 442-450.

Clarke, P., J. Leal, et al. (2008). "Estimating the Cost of Complications of Diabetes in Australia Using Administrative Health-Care Data." Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research **11**(2): 199-206.

Clarke, P. M., D. G. Fiebig, et al. (2008). "Optimal recall length in survey design." Journal of Health Economics **27**(5): 1275-1284.

Clarke, P. M., P. Glasziou, et al. (2010). "Event Rates, Hospital Utilization, and Costs Associated with Major Complications of Diabetes: A Multicountry Comparative Analysis." PLoS Med **7**(2): e1000236.

daCosta, M., J. C. Cappelleri, et al. (2011). "A Longitudinal Assessment of Painful Diabetic Peripheral Neuropathy on Health Status, Productivity, and Health Care Utilization and Cost." Pain Medicine **12**(1): 118-126.

Dixon, S., C. J. Currie, et al. (2000). "The cost of diabetes: time for a different approach?" Diabetic Medicine **17**(12): 820-822.

Duan, N. (1983). "Smearing Estimate: A Nonparametric Retransformation Method." Journal of the American Statistical Association **78**(383): 605-610.

Duan, N., W. G. Manning, Jr., et al. (1983). "A Comparison of Alternative Models for the Demand for Medical Care." Journal of Business & Economic Statistics **1**(2): 115-126.

Fetter, R. B. and J. L. Freeman (1986). "Diagnosis Related Groups: Product Line Management within Hospitals." The Academy of Management Review **11**(1): 41-54.

Fu, A. Z., Y. Qiu, et al. (2010). "Impact of concurrent macrovascular co-morbidities on healthcare utilization in patients with type 2 diabetes in Europe: a matched study." Diabetes, Obesity and Metabolism **12**(7): 631-637.

Gandjour, A., F. Kleinschmit, et al. (2002). "European comparison of costs and quality in the prevention of secondary complications in Type 2 diabetes mellitus (2000–2001)." Diabetic Medicine **19**(7): 594-601.

Gerdtham, U. G., P. Clarke, et al. (2009). "Estimating the cost of diabetes mellitus-related events from inpatient admissions in Sweden using administrative hospitalization data." Pharmacoeconomics **27**(1): 81-90.

Geue, C., J. Lewsey, et al. (2011). "Spoilt for choice: Implications of using alternative methods of costing hopital episode statistics." Health Economics **21**(10): 1201-1216.

Gilmer, T. P., P. J. Connor, et al. (2005). "Predictors of Health Care Costs in Adults With Diabetes." Diabetes Care **28**(1): 59-64.

Gold MR, S. J., Russell L, Weinstein MC. (1996). Cost Effectiveness in Health and Medicine. Oxford, Oxford University Press.

Hammer, M., M. Lammert, et al. (2009). "Costs of managing severe hypoglycaemia in three European countries." Journal of Medical Economics **12**(4): 281-290.

Jones, A. M. (2000). Health Econometrics. Handbook of Health Economics. Volume 1A.

Jönsson B. (2002), “Revealing the cost of Type II diabetes in Europe: CODE-2 Study”. Diabetologia; 45:S5-S12.

Laudicella, M., K. Rosen Olsen, et al. (2009). "What explains variation in the costs of treating patients in English obstetrics specialties?" Centre for Health Economics, University of York(Working Papers).

Levin, A., M. R. Chaudhry, et al. (2009). "Diabetes, kidney disease and cardiovascular disease patients. Assessing care of complex patients using outpatient testing and visits: additional metrics by which to evaluate health care system functioning." Nephrology Dialysis Transplantation **24**(9): 2714-2720.

Maciejewski, M. L. and C. Maynard (2004). "Diabetes-Related Utilization and Costs for Inpatient and Outpatient Services in the Veterans Administration." Diabetes Care **27**(suppl 2): b69-b73.

Manning, W. G. (1998). "The logged dependent variable, heteroscedasticity, and the retransformation problem." Journal of Health Economics **17**(3): 283-295.

Manning, W. G., N. Duan, et al. (1987). "Monte Carlo evidence on the choice between sample selection and two-part models." Journal of Econometrics **35**(1): 59-82.

Manning, W. G. and J. Mullahy (1990). "Estimating Log Models: To Transform or Not to Transform?" National Bureau of Economic Research Technical Working Paper Series **No. 246**.

Manning, W. G. and J. Mullahy (2001). "Estimating log models: to transform or not to transform?" Journal of Health Economics **20**(4): 461-494.

McCullagh, P. and J. Nelder (1989). Generalized Linear Model. Boca Raton, Chapman and Hall/CRC.
Mihaylova, B., A. Briggs, et al. (2010). "Review of statistical methods for analysing healthcare resources and costs." Health Economics: n/a-n/a.
Morgan, C. L., J. R. Peters, et al. (2010). "Estimated costs of acute hospital care for people with diabetes in the United Kingdom: a routine record linkage study in a large region*." Diabetic Medicine **27**(9): 1066-1073.
Nichols, G. A., B. Arondekar, et al. (2008). "Medical Care Costs One Year After Identification of Hyperglycemia Below the Threshold for Diabetes." Medical Care **46**(3): 287-292.
Norlund, A., J. Apelqvist, et al. (2001). "Cost of illness of adult diabetes mellitus underestimated if comorbidity is not considered." Journal of Internal Medicine **250**(1): 57-65.
O'Brien, J. A., L. A. Shomphe, et al. (1998). "Direct medical costs of complications resulting from type 2 diabetes in the U.S." Diabetes Care **21**(7): 1122-1128.
Office, T. S. (2008). NHS Costing Manual: 2007/08 edition D. o. Health. London.
Pagano E, B. M., Tediosi F, Garattini L. (1999). "Costs of diabetes. A methodological analysis of the literature." Pharmacoeconomics **15**: 583-595.
Palmer, G. (1996). "Casemix funding of hospitals: Objectives and objections." Health Care Analysis **4**(3): 185-193.
Ridyard, C. H. and D. A. Hughes (2010). "Methods for the Collection of Resource Use Data within Clinical Trials: A Systematic Review of Studies Funded by the UK Health Technology Assessment Program." Value in Health **13**(8): 867-872.
Schmitt-Koopmann, I., M. Schwenkglenks, et al. (2004). "Direct medical costs of type 2 diabetes and its complications in Switzerland." The European Journal of Public Health **14**(1): 3-9.
Schreyögg, J., T. Stargardt, et al. (2006). "Methods to determine reimbursement rates for diagnosis related groups (drg): a comparison of nine european countries." Health Care Management Science **9**: 215-223.
Tooze, J. A., G. K. Grunwald, et al. (2002). "Analysis of repeated measures data with clumping at zero." Statistical Methods in Medical Research **11**(4): 341-355.
UKPDS(33) (1998). "Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33)." The Lancet **352**(9131): 837-853.
UKPDS(34) (1998). "Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (34)." The Lancet.
UKPDS(38) (1998). "UK Prospective Diabetes Study 38: tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes." BMJ: 703-713.
Winkelmann, R. (2004). "Health care reform and the number of doctor visits - an econometric analysis." Journal of Applied Econometrics **19**(4): 455-472.
Zsolt, M. and P. Smith (2005). "The main methodological issues in costing health care services." Centre for Health Economics, University of York CHE Discussion Papers.

Chapter 4

USING HRQOL AS A PREDICTOR OF MORTALITY

Abstract

The aim of this chapter is to study the relationship between health related quality of life (HRQoL, e.g., EQ-5D) and mortality in patients with type-2 diabetes and to test whether HRQoL could be used as an independent predictor of mortality in patients with diabetes. This could help, for example, in the identification and management of high-risk patients using a HRQoL questionnaire as a tool to complement more expensive and detailed physical health assessments.

Out of the 5,102 patients participating in the UKPDS, 2260 (44.3%) died before the end of the study period (October 2007). In the follow up phase of the study, 3,380 patients answered at least one EQ-5D QoL questionnaire. The average length of time in the study was 17 years. The average time to the first QoL questionnaire administered since diagnosis was 10.1 years (SD:2.8).The average age at death was 68.7 (SD: 8.7).

A hazard model with time varying and time invariant covariates is estimated, controlling for demographic characteristics (gender, race, socio-economic status), biomarkers (albuminuria and creatinine), age and history of complications, together with behavioural variables (alcohol consumption, exercise and smoking status), in order to study the impact of self-reported quality of life on mortality. To test for the presence of frailty I compare predictions from a semi-parametric model with and without frailty. The semi-parametric approach used follows the Prentice and Gloeckler (1978) canonical model. It provides useful insights on possible omitted variables and it yields more plausible estimates and useful diagnosis than the more simple and common Cox proportional hazards model.

This study finds similar results to those published in the literature, namely that HRQoL is strongly and inversely associated with mortality. However, it also shows that if heterogeneity is taken into account this association increases on average. As time passes the slope of the individual hazard rises faster than the population hazard. This happens because over time the population is composed of more and more robust individuals as the frailer members exit the sample. This feature of duration dependence under heterogeneous frailty has the potential to bias estimates of the impact of time-varying characteristics, such as diabetes related complications and health related quality of life measures.

4.1 Introduction

The primary purpose of this chapter is to test whether measures of HRQoL based on index scores derived from responses to the EQ-5D questionnaire can be used as predictors of time to death in patients with type-2 diabetes. A number of studies have shown this to be the case in the context of diabetes as well as for other conditions (Section 4.2). A possible application of this research is to improve the identification of high risk patients in the management of diabetes care, alongside routine physical health assessments. Measures of HRQoL may provide subjective weighting of health problems that may not be captured with objective physical health assessments (Kiecolt-Glaser and Glaser 2002). These may encompass broader physical and mental factors, which can impact health by altering behaviour, adherence to treatment, or the function of the immune, endocrine and cardiovascular systems.

Herein the validity of this hypothesis is tested in the context of unobserved patient heterogeneity. This study's contribution is to determine more accurately the contribution of QoL to survival by both formally accounting for heterogeneity and by using time varying covariates, neither of which to the best of my knowledge have yet been explored in the literature. Thus the question I set myself to answer is whether HRQoL can truly be an independent predictor of survival or if it is more likely to simply be a catch-all variable. Repeated measurements of self-rated health may also help provide a dynamic evaluation of a patient, reflecting not only the initial level of health but also its trajectory. Fixed effects (FE) make it possible to control for all stable characteristics even if those characteristics cannot be measured. For event history analysis a FE version of Cox regression (partial likelihood) is available for data in which repeated events are observable for each individual (Yamaguchi 1986). But fixed-effects

Cox regression is not feasible when each individual experiences no more than one event, as is the case with death. Unobserved frailty of survivors may lead to dynamic selection, omitted variables correlated with life expectancy (LE) and QoL, as factors that influence LE may also influence tariffs. To circumvent this problem, I explore a different avenue and use a discrete time, semi-parametric maximum likelihood estimator of frailty, following the work of Prentice and Gloeckler (1978).

A review of the literature is presented in Section 4.2. There I look at both the justification for including QoL as an explanatory variable in outcomes like mortality and morbidity and I summarize the findings of peer reviewed studies. Section 4.3 discusses a number of methodological issues pertinent to this chapter. In particular it explains why unobserved heterogeneity is important even when unobservable characteristics are orthogonal to the covariates of interests and summarizes the possible consequences of ignoring it. Section 4.4 describes the data and the summary statistics in relation to the variables of interest. Section 4.5 illustrates the modelling strategy adopted for estimating survival in this context by outlining a discrete time duration model with gamma frailty. Section 4.6 reports on the results of the suggested model and compares it to those reported in the literature. Section 4.7 concludes and discusses the significance and limitations of the findings.

4.2 Literature review

The literature concerned with the association between self-reported measures of QoL and mortality, independent of "objective health status", is neither new nor limited to diabetes. Mossey and Shapiro's 1982 analysis of the Manitoba Longitudinal Study is considered one of the first studies of this kind. It showed that elderly Canadians' self-ratings of health were better predictors of seven-year survival than their medical records. That paper highlighted the possibility that self-reported questionnaires may unveil unreported symptoms that otherwise go untreated in the healthcare system. This finding seems to hold also in specific disease areas. For instance, index scores derived from the EQ-5D have also been shown to be predictive of short-term survival of terminal cancer patients (Park, Park et al. 2006) and of one-year survival after coronary revascularisation (Lenzen, Reimer et al. 2007).

The rationale put forward for this association is varied. Some suggest that depression contributes to disease and death through immune deregulation (Kiecolt-Glaser and Glaser 2002). Idler and Kasl (1991) suggested that self-assessments of health reflect a personal estimate of longevity, based not only on knowledge of the responder's own current health, but also on the knowledge of familial risk factors. Also, it is thought that poor perceptions of health may generate vicious cycles, leading to less engagement in preventive practices or self-care (non-adherence to recommendations, medication, and treatment). More broadly, there is a consensus that self-reported questionnaires on QoL could act as a relatively inexpensive proxy of other things researchers may not be able to observe or explain. Self-rated health is thus considered a more inclusive measure of health status and health risk factors than other covariates commonly used. Self-ratings surveys, which take only few minutes to obtain, thus appear to help in predicting

survival at the population level even when known health risk factors have been accounted for.

Idler and Benyamini (1997) reviewed 27 studies which showed that self-rated health predicted both short-term mortality (2–7 years) and long-term mortality (9–13 years). The papers included in their review show impressive consistency in findings despite differences in location, data collection and methods. Self-rated health is reported as an independent predictor of mortality in nearly all of the studies, despite the inclusion of numerous specific health status indicators and other relevant covariates known to predict mortality. In these studies QoL is measured by asking respondents to rate their overall health with the categories of excellent, good, fair, and poor, or some other variant. These studies have taken place in Sweden, Lithuania, Israel, Great Britain, the Netherlands, France, Poland, Hong Kong, Japan, Australia, Canada, and the United States. Follow-up periods ranged from 2 to 7 years in 20 of the studies, from 9 to 13 years in 6 studies, and 28 years in one study. In some studies, the outcome is treated as survival time; in others it is a binary variable for mortality. For most studies, odds ratios range from 1.5 to 3.0 for the extreme risk category. The odds of mortality for poor self-rated health often exceed those for smokers when they are reported in the same study. The consistency of the effects seems to show that the concept of self-rated health status is relatively insensitive to the semantic variations in the questions eliciting QoL.

Using PubMed online library as the main bibliographic data base, the following terms were searched: mortality, survival, hospitalization, EQ-5D and quality of life. Nine studies were included in this review; published between 2008 and 2012. Emphasis was given to patients with diabetes as a known pre-existing condition in order to have a benchmark for my findings. Two out of the nine studies published in the past 5 years

(Jerant, Tancredi et al. 2011; Cavrini, Broccoli et al. 2012) were however based on a generic elderly population. Most of these studies were part of larger trials testing the impact of different treatment interventions (e.g., the Fenofibrate Intervention and Event Lowering in Diabetes (FIELD), Translating Research Into Action for Diabetes (TRIAD), Zwolle Outpatient *Diabetes* project Integrating Available Care (ZODIAC)). When reviewing the most recent studies using QoL as a predictor of mortality and non-fatal complications in patients with diabetes, my search was limited to prospective studies using well defined cohorts instead of administrative data. In most countries, mortality records (based on hospital and clinic populations) are distorted by under-reporting diabetes as the underlying or contributory cause of death (McEwen et al., 2006). Table 4.1 summarizes the information collected, in alphabetical order.

All studies used a standard proportional hazard model. The majority of these left (implicitly or explicitly by stating the use of a Cox PH model) the baseline hazard unspecified while two modelled it parametrically: Landman et al. (2010) assumed a lognormal distribution (meaning that time follows a normal distribution) while Hayes (2008) assumed an exponential distribution (thus assuming a constant hazard over time). Not all studies reported the starting point of the survival evaluation. It can be inferred that some studies used age, diagnosis of diabetes or date of questionnaire response (in most cases the latter two corresponded to the same date) as starting points. Wennberg et al.(2012) had the longest follow up (a maximum of 14 years) although compared to the UKPDS it had a relatively low number of deaths (344 deaths compared to 2,260 deaths in the UKPDS, of which 1,057 had a preceding QoL measurement).

HRQoL in the studies reviewed was measured using two types of instruments: diseases specific questionnaires and generic questionnaires. The EQ-5D was the most widely used generic preference-based measure of health-related quality of life. Six studies (out of nine) used the EQ-5D questionnaire; two studies (based on the same data set[1]) used the RAND-36[2] (Kleefstra, Landman et al. 2008; Landman, van Hateren et al. 2010). Wennberg et al. (2012) used Self-rated health (SRH) which is defined by responses to a single question such as 'How do you rate your health?'.

Variables routinely included in these analyses included the traditional risk factors of blood pressure, cholesterol, glycated haemoglobin, smoking history, alcohol intake, weight, gender, age, and the respondent's disease history and current health status. Socioeconomic status (education, occupation and/or income level), marital status and ethnicity however were not always included (Hayes, Clarke et al. 2008; Clarke, Hayes et al. 2009; McEwen, Kim et al. 2009; Hayes, Clarke et al. 2011), yet these variables have been shown to independently predict mortality (Fiscella and Franks 1997).

All these studies indicate that HRQoL is an independent marker of mortality and this emphasizes the importance of looking beyond clinical parameters in both patients with Type-2 diabetes and the elderly. It is however possible that the significant associations between health measures and mortality reported in the literature may reflect confounding. The major methodological weakness of these papers in general is the inability to rule out residual confounding from other forms of co-morbidity or patient characteristics that were not assessed. None of these studies in fact evaluated or

[1] Kleefstra et al. (2008) followed patients for 6 years and Landman et al. (2010) had 3.6 years longer follow up.

[2] The RAND-36 contains nine health dimensions. These are usually sub-divided into a physical and mental score, ranging from 0 –100; higher scores indicate better HRQoL. The former has questions about ability to perform physical activities, the latter questions about patients' opinion about their health in general.

controlled for unobserved heterogeneity. More importantly none of these studies benefited from multiple rounds of QoL questionnaires. All the covariates included in the analyses were time-invariant. Covariates were either measured at baseline or as individual averages of available measures between randomisation and the time of the first administration of the questionnaire. Particularly disconcerting was the use of Quality Adjusted Life Years (QALYs) as the dependent variable in one of the specifications (Jerant, Tancredi et al. 2011) as QALYs are a function of survival time and QoL tariff, and therefore there must be an association by construction. Another possible limitation of these analyses is the poor response to the QoL questionnaires (ranging from 75% to 54%). The response rate in the UKPDS for instance varied from 90% in the first questionnaire to 74% in the last questionnaire.

Table 4.1: Overview of studies published on the link between self-reported quality of life and life expectancy

Author, year, study and country	Subjects, sample size and duration	Methods of analysis and instrument used	Covariates used	Outcome
Cavrini et al. (2012) Pianoro Study (Northern Italy)	9,644 elderly people with 5,256 completed questionnaires returned. Only one round of questionnaires collected. Five years observation period.	Multivariate Cox's proportional hazards models used to assess the effect of the EQ-5D score on 12 and 24 month mortality.	Gender, age, BMI, smoking status, activity levels, education, diabetes, morbidity levels. All entered as time invariant variables.	for a one unit increase in the index value, the risk of the occurrence of the events decreased by 69% and 58%, respectively, for mortality at 12 and 24 months.
Clarke et al.(2009) FIELD Study (Australia and New Zealand patients only)	7,348 respondents. One questionnaire. Median duration 10 years.	Multivariate Cox proportional hazard regression. Instrument used: EQ-5D.	Gender, smoking status, BMI, HbA1c, Systolic Blood Pressure, Total to LDL cholesterol ratio and diabetes duration. All entered as averages.	a 0.1 higher index score (derived from the UK algorithm) was associated with 14% (95% CI: 8-19%) lower rate of all-cause mortality.
Hayes et al. (2011) FIELD Study (Australia and New Zealand patients only)	7,348 patients. Two questionnaires. Median duration 8.3 years.	Exponential proportional hazards survival models for non-fatal diabetes complications and all-cause mortality. Time at risk was the time from the administration of the first EQ-5D questionnaire until trial closeout.	Gender, smoking status, BMI, HbA1c, Systolic Blood Pressure, Total to LDL cholesterol ratio. All entered as averages. EQ-5D was used as a change in measurement between the first and second questionnaire.	A 65-year-old patient at full health (utility = 1) can expect to live 2 years longer and achieve 6 more QALYs than a patient at average health (utility = 0.8), given similar clinical risk factors.

Hayes et al. (2008) FIELD study (Australia and New Zealand patients only)	7,348 patients. One questionnaire. Mean follow-up of 2.4 years after completing the EQ-5D questionnaire.	Cox proportional hazards model for fatal and nonfatal vascular events and complications of diabetes. Instrument used: EQ-5D visual analogues scale (VAS).	Gender, history of complications, A1C, BMI, smoking status, cholesterol, SBP.	A 10-point higher EQ-5D score was associated with a 6% (95% CI 1–11) lower risk of vascular events and a 22% (95% CI 15–28) lower risk of diabetes complications.
Jerant et al. (2011) MEPS (USA)	2000 to 2002 Medical Expenditures Panel Survey respondent data linked to the National Death Index through 2006.	Cox survival and area under the curve. Measures: EQ-5D summary index (EQ-5D); predicted EQ-5D (pEQ-5D) derived from the SF-12; SF-6D; and a single-item categorical SRH measure.	Estimates are presented with and without adjustment for socio-demographic characteristics (age, sex, race/ethnicity, education, and income).	A 0.1 points decline in the EQ-5D led to a HR=1.30 in the unadjusted model and 1.09 (1.00, 1.18) risk of mortality in the fully adjusted model (age, sex, race/ethnicity, income, education and baseline values of all study health measures).
Kleefstra et al. (2008) ZODIAC study (Netherlands)	857 patients, one questionnaire. Mortality assessed over 6 years.	Cox proportional hazards modelling. Instrument:RAND-36	Smoking, age, sex, diabetes duration, A1C, renal function, BMI, blood pressure, HDL cholesterol, and macrovascular complications.	The Physical Component Summary of the RAND-36 was inversely associated with mortality (hazard ratio [HR] 0.979 [95% CI 0.966–0.992]), as were the two separate RAND-36 dimensions.

Landman et al. (2010) ZODIAC study (Netherlands)	1,353 patients, one questionnaire. Mean follow-up of 9.6 years.	Cox proportional hazard model. HRQoL measure: RAND-36.	age, sex, smoking, duration of diabetes, serum creatinine, BMI, systolic blood pressure, total-to-HDL cholesterol ratio, macrovascular complications, use of statins, insulin use, and urinary albumin-to-creatinine ratio.	A 10-point-higher score on the Physical Component Score (PCS) and the Mental Component Score (MCS) decreased the risk for total mortality by 11% and 10%, respectively.
Mc Ewen et al. (2009) TRIAD Survey (multicenter, USA)	7,892 patients, one questionnaire, average length of follow-up: 3.7 years.	(i) Univariate proportional hazard model for unadjusted hazard rate ratios (HRs) for all-cause mortality, cardiovascular mortality, and non cardiovascular mortality. (ii) Multivariate Cox model. Instrument used: EQ-5D.	Charlson index, age, gender, income, duration of diabetes, BMI, race, smoking status and provider group/health plan cluster.	A 0.10 unit increase in the EQ-5D score was associated with an 8% decrease in mortality (HR 0.92 (95% confidence interval 0.87–0.96)).
Wennberg et al.(2012) EPIC study (Germany, Netherlands and Sweeden)	3,257 individuals. One questionnaire. Mean follow up 8.6 years.	Cox proportional hazards model. Instrument: Self-rated health (SRH).	Age, centre, sex, educational level, body mass index, physical inactivity, smoking, insulin treatment, hypertension, hyperlipidaemia and history of myocardial infarction, stroke or cancer.	The fully adjusted HR for low SRH was 1.38 (95% CI 1.10 to 1.73).

4.3 Methodological issues

Epidemiological analysis of prognoses of T2DM, as is the case with many other chronic conditions, presents considerable methodological problems. This section will summarize the different strategies available to model the frailty of survivors.

It is important to note that in duration models unobserved heterogeneity always have the potential to affect the estimated coefficients as well as the baseline hazard, even if unobservable variables are orthogonal to the covariates. This is because as time proceeds, there is, by construction, sample selection. For instance, one could imagine two types of patient, each with different hazards. Assume there exists a variable X that helps us identify the two different groups (e.g., male and female) by telling us the magnitude of the difference between these two groups. If we exclude X from our model, over time, the observations with the higher hazards (representing patients who are more "frail") will fail and thus exit the sample and those with a lower hazard (less frail) will remain. Thus the estimated hazard will appear to decline over time. This leads to a discrepancy between the initial distribution and the final distribution of survivors in the sample. This spurious duration dependence will always make duration dependence appear more negative than it really is. The result is that flat hazards will look as if they are declining and rising hazards will appear flat or even non-monotonic. This issue of dynamic selection in the literature is often illustrated as the mover-stayer problem. Lancaster (1990) shows that when the initial distribution is different from the final distributions of survivors in the sample, the coefficients on all observed variables are biased towards zero.

This discussion raises the issue of identification. The question of interest thus becomes whether we can disentangle true versus spurious duration dependence. A number of patient-specific omitted variables may be correlated with survivorship. These include genetic variations, heritability, etc. Moreover, the same factors that influence life expectancy may also influence self-reported QoL, creating endogeneity. These can be economically relevant factors such as income or employment status, situational variables denoting health and social relations, socio demographic indicators (gender, age, education), or other genetic or personality traits. These examples raise the issue that whatever one may be picking up as the effect of QoL on survival may be spurious and the result of other exogenous variables that happen to affect both longevity and self-reported well being.

Standard approaches

The analysis of survival data can take one of three forms: non-parametric, semi-parametric and parametric. In the case of a non-parametric analysis, no assumption is made about the functional form of the survivor function. In this setting, two approaches are commonly used: The Kaplan-Meier (1958) and the Nelson-Aalen estimator (1978). The latter is deemed to have better small sample properties. Both estimate the probability of surviving past time t.

$$\hat{S}(t) = \prod_{j|t_j \leq t} \frac{n_j - d_j}{n_j} \qquad (1)$$

where n_j is the number of individuals at risk at time t_j and d_j is the number of failures at time t_j. Eq. 1 is then an estimate of the running product of the probabilities of surviving at different time intervals. The drawback of the non-parametric approach is that it is not generalizable. The effects of covariates are not modelled and comparison of survival

can only be done across the values of categorical covariates. Furthermore, the Kaplan-Meier hazard assumes that the sample is homogeneous, i.e., that there is no heterogeneity which depends on either observable or unobservable factors. However, one expects that the characteristics causing a lower hazard will be more concentrated among individuals who stayed longer in the sample and are thus older and more frail.

The Cox model (1972) is one of the most popular semi-parametric models. Although no survival function is specified ($h_0(t)$, the baseline hazard can be left unspecified and estimated non-parametrically), the effects of the covariates (β_x) are parameterized to multiplicatively shift the baseline hazard function as follows.

$$h(t|\boldsymbol{x_j}) = h_0(t)\exp\,(x_j\beta_x) \qquad (2)$$

Rather than directly modelling the length of the interval, the dependent variable in the Cox regression is the hazard or instantaneous likelihood of event occurrence. $\boldsymbol{x}_i$ is a column vector of predictor variables that may vary across individuals. Hazard ratios are interpreted in the same way as odd ratios in logistic regression. The central assumption here is that the hazard over time, regardless of its form, is the same for everyone (Cleves, Gould et al. 2010). The downside of this model is loss of efficiency vis a vis fully parameterized models. If we knew the functional $h_0(t)$ form we could do a better job of estimating β_x (Lin and Wei 1989). The Cox regression model also requires considerable computational power when estimating time varying covariates, although it is precisely the time dependent covariates that make a duration model especially useful (Meyer 1990).

Parametric estimation helps to obtain more efficient estimates of the coefficients of interest than the Cox model and it also allows us to obtain an estimate of the underlying baseline hazard. In other words, parameterization allows us to know the

absolute risk as well as the relative risk. For instance, if the mortality rate were to double for individuals who undergo a 50% decrement in utility, then we might also want to know what reference value this doubling refers to, and check whether the proportionality assumption is valid or not. There are six standard parametric survival models: exponential, Weibull, Gompertz, lognormal, loglogistic and generalized gamma. The first three distributions have monotonic hazards (they increase or decrease over time, or stay constant, in the case of the exponential distribution). The last three distributions have hazard functions with turning points. It is thus not possible to have proportional hazards (PH) with the lognormal, loglogistic or generalized gamma as these can only be expressed in the accelerated failure-time metric[3].

While parametric models allow us to obtain values for the baseline hazard and smooth estimates of the survival functions for any combination of the covariate values and, when needed, time-dependent hazard rates (HRs), misspecification of duration dependence may produce misleading results for β_x. Unfortunately, it is not possible to compare a Cox model with a parametric model by model fit because the former is estimated by maximum *partial* likelihood and thus does not generate information criteria (Akaike or Bayes Schwarz). Model selection criteria are thus determined by diagnostics other than those dictated by model fit[4].

[3] Accelerated failure-time models are used in other disciplines. In medical research interest lies into the actual risk process and how this changes with covariates. Hence attention is given to the PH metric rather than predictions that give a prominent role to the analysis of time and the actual failure times.

[4] When parametric models are nested, the likelihood-ratio or Wald test can be used to discriminate between them. This can be done for Weibull versus exponential or Gamma versus Weibull or lognormal. When models are not nested, however, these tests are inappropriate, and the task of discriminating between models becomes more difficult.

Heterogeneity and Frailty

A number of omitted variables may exist that are both patient specific and correlated to the outcome of interest: genetic variation and heritability are just a few examples that may be linked to survivorship. Failure to address duration dependence leads to under-estimation of standard errors and p-values (Aalen 1988). It is possible to separate the duration dependence due to unobserved heterogeneity from the duration dependence due to sample selection (also known as state dependence). The way this is often done is by assuming a parametric distribution for the frailty. Ideally one might want to estimate a different hazard for each patient to account for fixed effects (FE). Using dummy variables to model all subjects however is not only unviable practically given the high number of indicator variables that would have to be added, it would also introduce bias (away from zero) by incorporating so many incidental parameters.

Rather than employing dummies or first differencing (impossible here since the functional form is not linear) one can obtain FE in a way similar to logistic and Poisson regression, where the coefficients for the dummy variables are actually not estimates but are eliminated from the likelihood function. Fixed effects regression requires that at least some of the individuals in the sample experience more than one event so that the within-individual comparisons are possible. The method therefore cannot be applied to non-repeatable events such as deaths. Nevertheless, under certain conditions it may be possible to do a FE analysis for non-repeatable events by treating time as discrete and applying conditional logistic regression. Maclure (1991) calls this *case-crossover analysis*.

Due to computation problems in estimating FE, to evaluate the contribution of hidden variations to observed hazards, random effects (RE), often referred to as *frailties,* are used to deal with unobserved heterogeneity between clusters. These might be

individual-specific or group-specific (for example, different hospitals may have different frailties). The idea in the context of this analysis is that some individuals are frailer than others and hence are more likely to experience death. Vaupel et al. (1979) for example suggest using such models when individual variations in survival chances due to advanced age would lead to biased estimates of coefficient in Cox-type models of hazard rates. The idea here is that individual variations in survival probabilities due to observables (e.g., advanced age) may lead to hidden variations in the hazard rates as health decreases with age and self-reported QoL decreases with age.

Multiplicative Frailty

A frailty is a latent random effect that enters multiplicatively into the hazard function. For the jth subject in the ith group (*i* and *j* can be made to coincide) the hazard is:

$$h_{ij}(t) = h_0(t)\alpha_i \exp(x_{ij}\beta) \qquad (3)$$

where α_i is the group-level frailty. The frailties are unobserved positive quantities and are assumed to have mean=1 (this normalization is useful for identification reasons) and a finite variance estimated from the data (I call this θ). Just as in RE linear models the idea is that observations within patients are correlated because they share the same frailty. The extent of the correlation is measured by the variance of the frailty: when the variance of the frailty is zero, the frailty model reduces to a standard hazard model.

This approach is multiplicative because unobserved differences between observations are introduced via a multiplicative scaling factor `alpha'. The main assumption of this *multiplicative frailty* model is that α_i *is distributed independently of* x_{ij} *and t.* The random effects (RE) α_i is not observed/observable but is assumed to have a

distribution. In order for the model to be identifiable the frailty has to be fixed to some parametric distribution. Maximum likelihood is then used to estimate the parameters of the structural duration distribution and the distribution of unobservables.

The distribution on the unobservable is however ad hoc. While a theory may offer some guidance on the appropriate functional form for observable variables, it rarely offers guidance on the functional form of the distribution of unobservables. The choice of a particular distribution is usually justified on the grounds of familiarity, ease of manipulation, and considerations of computational cost (Jenkins 2008). For mathematical tractability, the distributional choice for α is often limited to either the gamma distribution or the inverse-Gaussian distribution. The output of the estimation remains unchanged from the non-frailty version, except for the additional estimation of θ (the variance of the frailty).

The use of frailty models has its limitations. In order for the model to be identifiable the frailty has to be fixed to some parametric distribution. Common criticisms, albeit general to any other model, are that incorrect parametric specification can bias the estimates of regression coefficients. Also one should note that under different functional forms, one is expected to obtain different values for the variance of the frailty. While the variance acts as a measure of the degree of heterogeneity in the population, there is in fact no such absolute measure. The “amount of heterogeneity” measured in a statistical analysis of survival data is determined by identifiability conditions. Another limitation of frailty models is that they maintain the proportional structure of the hazard. This, in itself, is not a direct criticism of frailty but a limitation of the simplicity of Cox models.

Non-parametric frailty

Heckman and Singer (1984) have shown that different distribution functions of duration can explain the same data and yet be misleading either at the individual or at the aggregate level and are thus a wrong representation of heterogeneity[5]. Because of this instability of parameter estimates depending on the type of assumption made about the distribution of frailty, Heckman and Singer (1984) proposed a solution using a non-parametric maximum likelihood estimator (NPMLE) of the distribution of frailty. Frailty is then represented as set of 'mass points' together with the probability of each subject being at each mass point. Individuals belong to one of a number of different types but membership to each class is unobserved. This feature is incorporated in the Heckman and Singer model by allowing the intercept to vary between the groups.

Assuming for simplicity only two groups, if $\alpha_2 > \alpha_1$ then type-2 are fast exiters relative to type 1 people, other things being equal. This specification is a form of latent class model where the process describing time-to-event differs between the classes (groups) within the population. There are M latent classes and the aim is to find the probabilities of a person being located at each mass point. The number of mass points to choose is not obvious *a priori* so in practice, researchers have often used a small number (M=2 or M=3). One need not assume the number of classes, but can always test whether the addition of an extra class adds information. Trussell and Richards (1985) studied whether models estimated using the Heckman-Singer technique were sensitive to the

[5] Heckman and Singer use as an example the duration of a single unemployment spell. In the literature there are two models. One model assumes that at the onset all individuals that become unemployed are the same but as time progresses those who stay in unemployment longer because of some random matching process will end up being less likely to gain employment as they would have lost skills, become discouraged, etc. The second model assumes that at the onset people are different and as such they will exit the unemployment pool at different rates corresponding to their unobserved heterogeneity. Although the predictions of these two models are the same: a person who is unemployed for a long time is more likely to stay so. The mechanism is quite different, one is duration dependence and the other is selection.

choice of baseline hazard. They found that results are more sensitive to the choice of the baseline hazard than to the choice of the distribution of α_i.

Alternatives: Double parametric models and mixed models (parametric frailty and semi-parametric hazards)

The Heckman and Singer model is notoriously difficult to implement in that it often leads to non-convergence and needs substantial computer power. One alternative is to use a fully parametric model, one that is parametric in both the frailty and the proportional hazards. These models have the advantage of making predictions easy. With it one can generate a survival curve that averages over the distribution of all the explanatory variables[6]. But, as discussed in the previous section, there are shortcomings to parametric models, namely lack of flexibility.

It seems that a sensible approach in this case, especially when one does not fully understand the data generating process, is to use a duration model, which can be less parametric about the shape of the distribution and still allow censoring. Thus, the estimation approach used here is an extension of Prentice and Gloeckler (1978) which is discussed extensively in Meyer (1990). In it the shape of the hazard is non-parametrically estimated. This approach has several advantages. First, the estimates are parameters of a hazard model and thus retain an easy interpretation. Second, the probabilities of surviving each period are constrained to lie between 0 and 1. Third, it

[6] It is questionable whether one might actually want to do that. In fact it is hard to ascertain, what is the distribution of the explanatory variable in a survival analysis? The distribution at t=0, or at each individual time point. In some sense the latter seems more attractive, but changes in the survival curve then also represent changes in the distribution of the explanatory variables in the at risk population. It may infact be more meaningsul to illustrate a representative patient rather than the population average, which is precisely what is possible in semi-parametric models (Cox model and the Prentice and Gloeckler model to be illustrated).

allows for the presence of unobservable differences among people. Finally, it is relatively easy to test for heterogeneity and to estimate it parametrically[7].

Consequences of ignoring frailty

An important question to ask before fully describing the methodological approach undertaken (Section 4.5) involves the consequences of ignoring unobserved heterogeneity. The theoretical literature has suggested the following consequences, typically derived with reference to a continuous time PH model:

1. The non-frailty model will over-estimate the degree of negative duration dependence in the (true) baseline hazard, and will under-estimate the degree of positive duration dependence. The term *duration dependence* refers to the extent to which the conditional hazards of the events of interest are rising or falling over time. In the negative duration dependence case, observations with high α values fail faster, other things equal, so the survivors at any given survival time are increasingly composed of observations with relatively low α values and hence lower hazard rates.
2. The proportionate effect of a given regressor on the hazard rate is no longer constant and independent of survival time but is driven by selection and its relation to the explanatory variables. The distribution of frailty changes due to the selection effect as we allow for time-varying covariates.

The empirical literature has generally confirmed these consequences. There has also been discussion of the magnitude of the effects and how serious the biases are in practice (Royston and Lambert 2011). Verdicts have been contingent on the choice of

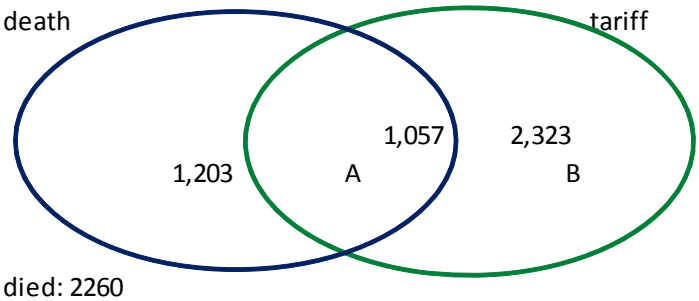
[7] The distribution of the heterogeneity component can be also estimated non-parametrically a priori under this framework.

shape of the non-frailty hazard function and the choice of the distribution for the unobserved heterogeneity (Aalen 1988). The consensus appears to be that if a fully flexible specification for the baseline hazard function is used, the magnitude of the biases relative to the 'true' model are diminished (Heckman and Singer 1984; Meyer 1990).

4.4 The data

In the UKPDS Post Trial Monitoring Study, the EQ-5D was administered a maximum of seven times over the 1997-2007 period to 3,380 patients, 1,057 of whom died. Time is defined from the time of diagnosis of diabetes. Participants were followed until death, migration, withdrawal or end of follow-up period. The outcome of interest here is death from any cause. Date of death was an outcome of the UKPDS and collected for all patients. The EQ-5D questionnaire was administered to UKPDS participants at a mean (SD) time of 10.7 (3.11) years after diagnosis and randomisation to therapy. Those who answered the first questionnaire would have done so at least 5 years after entering the trial for the late comers recruited in 1991 or at most 18 years for those recruited in 1978.

Figure 4.1: Information Sample



died: 2260
have a tariff: 3380
A: have a tariff and died
B: have a tariff and are censored

Figure 4.2: Kaplan-Meier survival estimate since diabetes diagnosis (over 30 years)

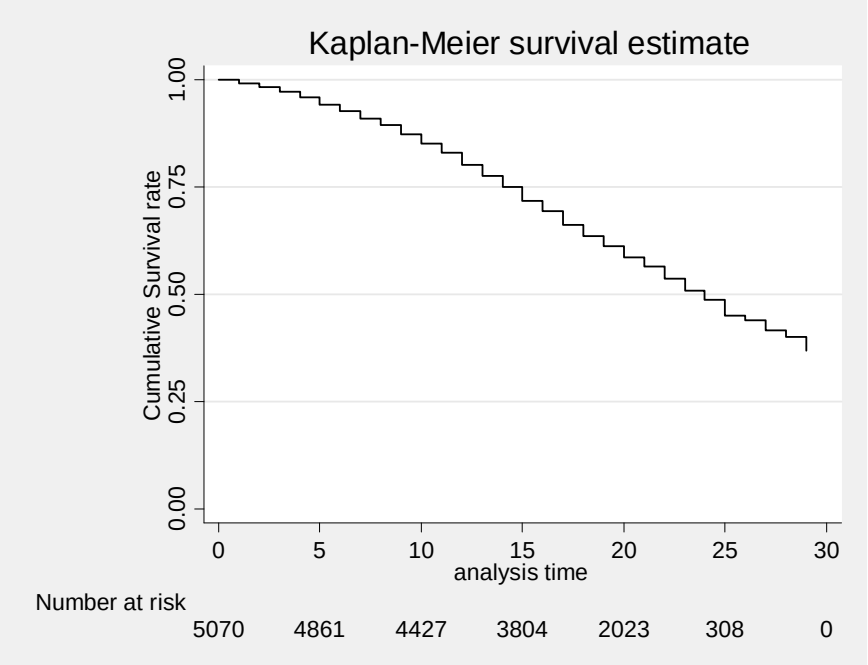


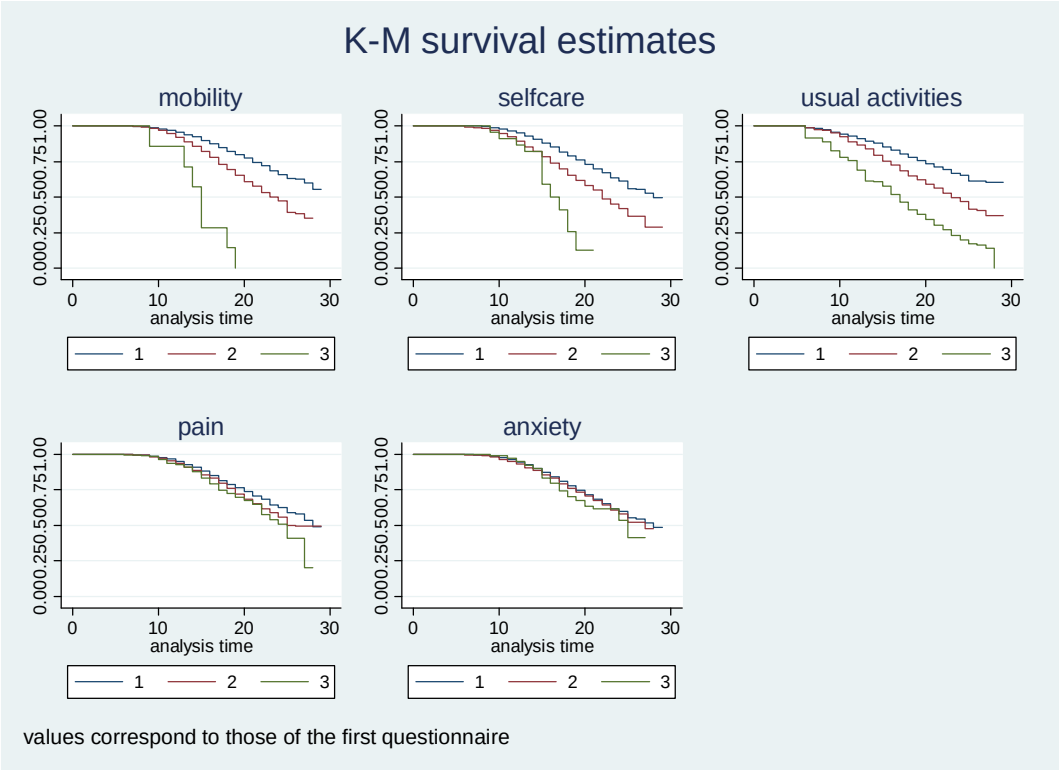
Table 4.2: Causes of death among those with and without tariffs

	Total	**No tariffs**	**At least one tariff**
Major vascular deaths*	1,098	606 (55.6%)	492 (44.8%)
Other deaths	1,162	597 (51.4%)	565 (48.6%)
Total	**2,260**	**1,203 (53.3%)**	**1,057 (46.7%)**

*Stroke, CHD, MI

To visually examine the degree to which risk varies across ranges of the EQ-5D components, survival was analysed according to patients' response level in each of the five dimensions without adjustment for covariates (Figure 4.3).

Figure 4.3: Graphical assessment of the proportional hazard assumption



Those with severe problems in any dimension (=3), present lower survival estimates than those with no problems at all (=1). Dimensions related to perceived physical capabilities (mobility, self-care and, usual activities) have more power in disentangling variations across levels (no problem=1, some problem=2, extreme problems=3) than perceived psychological dimensions (pain and anxiety). It is worth noting however that only 37 patients reported mobility=3 in the EQ-5D questionnaire and 16 of them died in our sample. There are many non-parametric tests available to gauge the statistical significance of these differences. I use the log-rank test to compare the overall survival across the three groups in each of the categories. The null hypothesis is no difference in survival between the three groups. The column "*Events observed*" in Table 4.3 refers to the number of failures observed while the "*Events expected"* column

refers to the number of events that would be expected if the groups shared the same survivor function.

Table 4.3: Difference in the levels of EQ-5D components in explaining mortality

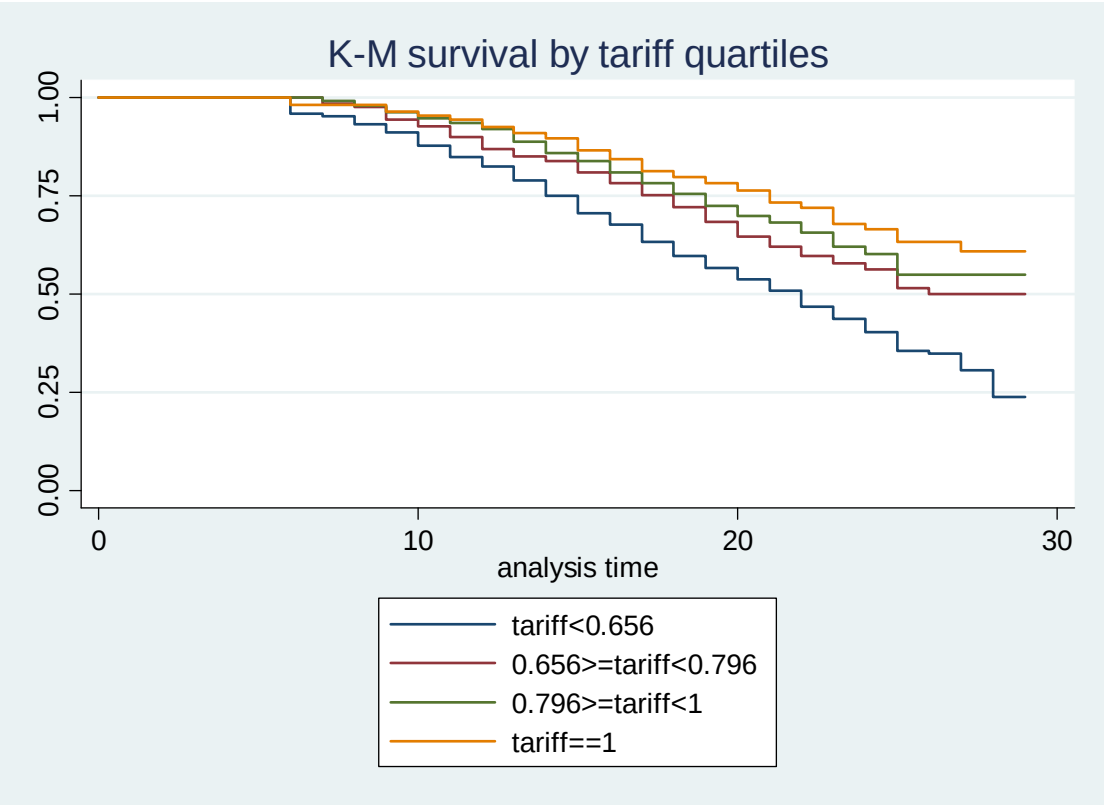
dimension	level	deaths observed	deaths expected	Log rank test chi2(2)	Wilcoxon chi2(2)
mobility	1	444	606.21	145.81**	128.83**
	2	595	446.59		
	3	18	4.2		
self care	1	807	908.66	129.64**	95.34**
	2	211	137.46		
	3	39	10.87		
usual activities	1	514	687.88	201.11**	163.65**
	2	415	318.46		
	3	128	50.66		
pain	1	418	464.73	10.29**	7.46*
	2	532	502.65		
	3	107	89.63		
anxiety	1	652	695.94	10.45**	6.99*
	2	356	324.5		
	3	49	36.57		

**P-value<0.01, * P-value<0.05

The log rank test depends on the proportionality assumption of the hazards. Because the hazards seem to vary in ways other than proportionally, the Wilcoxon test was also used. The Wilcoxon test statistic is constructed by weighting the contribution of each failure time to the overall test statistic by the number of subjects at risk. Thus, it gives heavier weights to earlier failure times when the number at risk is higher. As a result this test is susceptible to differences in the censoring pattern of the groups (StataCorp 2011). If the survival curves cross, as it is the case in Figure 4.3, then neither the log-rank nor the Wilcoxon tests are helpful. Because the crossings in mobility, self-care and pain happen at the beginning, rather than the middle, of the time under observation, this could represent random sampling, and the assumption of proportional hazards could still be valid.

In order to visually examine the degree to which risk varies across ranges of the aggregated index score, estimates of the survival function were plotted for the following groups (mean of each group reported in brackets): index score < 0.656 (0.325); index score >=0.656 and index score < 0.796 (0.71); index score >=0.796 and index score < 1(0.821); index score = 1.00 (1.00). Each group represents an approximate quartile of the data. Standard log-rank methods without adjustment for covariates were used to test for significant differences between these groups for each specified endpoint. There is a clear difference in survival across groups.

Figure 4.4: Kaplan–Meier quartiles of tariff over 30 years



Quartile	Mean tariff	Deaths observed	Deaths Expected
tariff<0.656	0.325	374	250.22
0.656>=tariff<0.796	0.710	252	247.16
0.796>=tariff<1	0.821	193	216.25
tariff==1	1.000	238	343.37
Total		1057	1057

Log Rank chi2(3) 100.91**; Wilcoxon chi2(3) 83.8**

** P-value<0.01

The proportional hazards assumption for the tariff as a whole was examined using a test based on Schoenfeld residuals. After using a simple specification where the tariff is the only variable included, a smooth function of time is fitted to the residuals. Under the null hypothesis of proportional hazards, we expect the relationship between the scaled residuals and time to have zero slope. This has been shown to be equivalent to testing that the log hazard ratio function is constant over time (Schoenfeld 1982). A rejection of the null hypothesis indicates a deviation from the proportional-hazards assumption (PH). The test did not indicate a violation of PH with chi2(1)=1.02, Prob>chi2=0.312 (For a graphical representation see Appendix A4.1). Similar results are obtained after controlling for the whole set of patient characteristics available in the UKPDS data.

We have three sets of covariates available: those that are strictly time invariant (e.g, gender and ethnicity), those that are time varying but were not collected throughout the follow up period (biomarkers, exercise and smoking) or were collected only once (e.g., socioeconomic status), and those that were regularly updated and measured at different points across the follow up period (notably, tariffs, age and non-fatal complications). Over the course of the entire trial 2,505 patients out of 5,102 incurred a non fatal event. Of these 41.77% were macrovascular and 58.23% were micro-vascular. Over the ten year follow-up the average number micro-vascular events is 4.5% per annum and the mean number of macro-vascular events is 2.1%.

The average tariff is slightly higher than the one reported in Chapter 2 (0.69 vs 0.70). This is because whenever a participant did not complete a survey round the value of his or her last available tariff was carried forward until the next survey completed or until death or exit from the sample. The fact that the average tariff is slightly higher suggests

that unanswered questionnaires are followed by questionnaires with lower valuations and that the last valuation before exiting the sample is also lower than any previous one. Because questionnaires are administered on the anniversary of therapy decision, if we consider deaths occurring strictly within a year of a questionnaire response, we have 159 deaths. If, however, we allow for the last QoL measurement available to predict death, we could include all 1,057 deaths in the estimation. The mean time between QoL measures and death is 3.83 (SD 2.91) years. It is important to note that this could potentially add noise to the estimation but at the same time, it allows us to use the whole sample of questionnaire responders.

The study population included 1,984 men (58.7%). Mean age at first EQ-5D questionnaire was 62 while the average age in the sample was 67(SD 9.4).

Values of all risk factors in Table 4.4 are calculated as an across patients average of their last measurements at the time of first administration of the EQ-5D. Each patient's risk factor (smoking status, exercise, albuminuria and creatinine) is carried forward until death or censoring. Risk factors such as HbA1c, SBP and BMI are known to influence complications (Stratton, Adler et al. 2000). The inclusion of biomarkers is, however, problematic in the estimation of survival whenever we also control for complications because they affect mortality by increasing the likelihood of complications. It is not surprising then that most of the biomarkers tested (blood pressure, BMI, lipoproteins and HbA1c) were not statistically significant when occurrence of complications was also included.

Table 4.4: Characteristics of patients completing the EQ-5D questionnaire

Variables in the model N=(3,380)	Mean (SD)
Tariff	0.70 (0.30)
Age	67(9.40)
Microvascular	4.5%
Macrovascular	2.1%
Male	58.7%
Caucasian	82.7%
Black	6.6%
Asian	10.7%
Skilled	42.8%
Unskilled	57.2%
Smoker	64.3%
Sedentary	38.4%
Active (occasionally)	53.1%
Fit	8.5%
Albuminuria (mg/h)	96.83(413.9)
Creatinine (μmol/l)	94. 4(35)

4.5 The Models

Herein two models are described: (1) the Prentice-Gloeckler (1978) model; and (2) the Prentice-Gloeckler (1978) model, incorporating a gamma mixture distribution to summarize unobserved individual heterogeneity, as proposed by Meyer (1990).

As an example suppose there are individuals i = 1,...,N, who each enter a state (i.e., they develop diabetes) at time t = 0 and are observed for k_i time periods, at which point each person either remains in the state (censored duration data) or leaves the state (complete duration data). Consider a proportional hazards function as defined in equation (2):

$$h_{it}(t) = h_0(t)\exp\left(\boldsymbol{x_{it}\beta}\right)$$

where $h_0(t)$ is the baseline hazard at time t, which is unknown, x_{it} is a vector of time-dependent explanatory variables (covariates) for individual i, and β is a vector of parameters which is unknown. This is a time discrete model in that the covariates remain constant between t and t+1. The probability of being observed in the next period can be written as follows:

$$P[T_i \geq t+1 | T_i \geq t] = \exp\left(-\exp\left(x_i(t)'\beta + \gamma(t)\right)\right)$$

where $\gamma(t) = ln\left\{\int_t^{t+1} h_0(u)du\right\}$

The likelihood function can be written as:

$$l(\gamma,\beta) = \prod_{i=1}^{N}\left[1 - \exp\left\{-\exp[\gamma(k_i) + x_i(k_i)'\beta]^{\delta_i}\right\} \times \prod_{t=1}^{k_i-1} \exp\left\{\exp[\gamma(t) + x_i(t)'\beta]\right\}\right]$$

where $\gamma = [\gamma(0)\gamma(1) \ldots \gamma(T-1)]$; $\delta_i = 1$ if $T_i \leq C_i$ where C_i is the censoring time and zero otherwise. $k_i = \min\left(int(T_i, C_i)\right)$.

The corresponding log-likelihood function and the equation we actually estimate is:

$$L(\gamma,\beta) = \sum_{i=1}^{N} \delta_i \log[1 - \exp\{-\exp(\gamma(k_i) + x_i(k_i)'\beta)\}] - \sum_{t=1}^{k_i-1} \exp[\gamma(t) + x_i(t)'\beta]$$

(4)

The conventional approach assumes a parametric form for the baseline hazard. This is equivalent to restricting the value of γ. The Prentice and Gloeckler approach is similar to Cox's model (Cox (1972)). Both estimators make no assumptions about the baseline hazard.

Unobserved heterogeneity is assumed to take a multiplicative form as in equation (3) where α_i is a random variable and is assumed to be distributed independently of x and t.

In principle, any distribution with positive support and finite variance is a suitable candidate to represent the frailty distribution. For tractability reasons, however, the

choice of distribution is typically limited to those that provide a closed form expression for the frailty survivor function, avoiding numerical integration. For the discrete time PH model, the Gamma distribution has been the most popular (Jenkins 2008) and is the one used in this study; other distributions could in principle be used. If α_i is distributed gamma with mean one (a normalization) and variance σ^2, then the log-likelihood becomes

$$L(\gamma,\beta,\sigma^2)=\sum_{i=1}^{N} log\left\{\left[1+\sigma^2\sum_{t=0}^{k_i-1} exp\{\gamma(t)+x_i(t)'\beta\}\right]^{-\sigma^{-2}}\right.$$
$$\left.-\delta_i\left[1+\sigma^2\sum_{t=0}^{k_i} exp\{\gamma(t)+x_i(t)'\beta\}\right]^{-\sigma^{-2}}\right\} \quad (5)$$

As σ^2 tends to zero, the population and individual hazard functions coincide.

If the distribution of α_i is unknown, γ and β can be also consistently estimated using the Heckman and Singer (1984) approach. α_i is modelled as a discrete distribution with a finite number of mass points. In practice, I have found the computation of that model problematic. Meyer (1990) discusses some reasons why this may be the case. In the section below, estimates without heterogeneity and with gamma distributed heterogeneity are reported and compared.

Predictions and the baseline hazard function

Just as in the Cox PH model one needs to either parameterize time or use a more flexible alternative such as time dummies in order to obtain the baseline hazard. Recall that time herein is defined as time since therapy decision. My aim here is to be agnostic and fully exploit the non-parametric baseline hazard. Therefore, time is modelled as a

piece-wise constant $\gamma(t) = D1 - D4$ (corresponding to years 1 to 4), $D5 - D9$ (years5 to 9), $D10 - D14$ (years 10 to 14), $D15 - D19$ (years 15 to 19), $D20 - D24$ (years 20 to 24), $D25 - D29$ (years 25 to 29).

For the latest entrant in the study, five years elapsed before he/she answered the first questionnaire. This means $\gamma(t)$ – the function of time – will not be identified between time 0 and 5. However, the survivorship function can be defined for these periods as this is anchored to 1 at time=0.

A matrix score is used to calculate the median survival for a representative patient at different quartiles of the EQ-5D tariff. To ensure that the values attached to the tariff and the other characteristics of the representative patient reflect those of the individual, rather than sample changes over time, all values correspond to those of the patient at first questionnaire. Having generated the predicted hazard and survivor probabilities, I use the latter to construct partial age-specific life tables for these representative patients. These are constructed using the rates of death obtained from the survival functions over a 20 year period. The number of person years was then calculated $L(t) = \frac{l(t)+l(t+1)}{2}$ where $l(t)$ is the estimated number of people alive. The estimated average life years lived is then calculated as $LE(t) = \sum L(t)\ /l(0)$, where $l(0)$ is the starting number of patients in the cohort (a schematic representation of this exercise and the assumptions made are presented in Appendix A4.5).

4.6 Results

In Table 4.7, the maximum likelihood (ML) estimates of the two discrete time proportional hazards regression models discussed above are presented: one without heterogeneity (Equation 4) and one with a gamma mixture distribution (Equation 5) to summarize unobserved individual heterogeneity ('frailty'). Covariates include time-varying and invariant regressors.

I find that a 0.1 increment in index score is associated with a 4.8% (95% CI 2.8%-6.7%) lower risk of mortality. If heterogeneity is taken into account, a 0.1 increment in index score is associated with a 5.4% (95% CI 3.1%-7.5%) lower risk of mortality. Overall, higher hazard rates (i.e., shorter survival times) are positively associated with age, males, micro and macrovascular events, lower socioeconomic status, and a number of important risk factors such as lack of physical activity, albuminuria and creatinine[8] at conventional levels of statistical significance. To facilitate interpretation albuminuria and creatinine have been standardized using the mean and standard errors preceding the first questionnaire in order to maintain representativeness within samples (a standardization that took into account the mean and standard deviation across all time periods will weigh more those who live longer). The time dummies reported at the bottom of Table 4.7 show that there is positive duration dependence and that this is roughly monotonic. An F-test conducted to see if these might be the same rejects the hypothesis (chi2(4) = 11.94; Prob > chi2=0.0178).

The σ^2 value reported in the output is the estimate variance of the frailty and this is equal to 0.4598. This means that the average difference in hazards across

[8] Herein Systolic blood pressure, BMI, high and low density lipoprotein and glycosylated haemoglobin (Hba1C) were tested as possible covariates in addition to albuminuria and creatinine but were not found to be significant.

individuals is 45.9% independently of the value of their covariates. Note that the frailty model is preferred to the reference non-frailty model according to the relevant likelihood ratio test. The test is a 'boundary' test that takes account of the fact that the null distribution is not the usual chi-squared (d.f.=1) but is rather a 50:50 mixture of a chi-squared(d.f. = 0) variate (which is a point mass at zero) and chi-squared(d.f. = 1). Hence the reference to 'chibar2(01)' in the output (Gutierrez, Carter et al. 2001). The P-value =0.005 in this case. Appendix A4.2 reports the untransformed coefficients corresponding to Table 4.7, i.e. β rather than $\exp(\beta)$.

The estimated decrement in the likelihood of survival from an increase in the tariff is larger in magnitude in the frailty model than in the reference model. The difference is not statistically significant in the model with all the covariates available from the study. However, if we take away covariates and use tariffs as the sole explanatory variable, we find the difference in the two models to be stronger and statistically significant (Table 4.5).

Figure 4.5 illustrates the distribution of the frailty with mean one and variance 0.459. The quartiles of the frailty gamma distribution are shown in Table 4.6. These are 0.502, 0.852 and 1.338. Computing the ratios Q1/Q2 and Q3/Q2, we see that individuals with frailty at Q1 have 41.1% lower risk, and individuals with frailty at Q3 have 57.19% higher risk, than those with median frailty. Clearly, unobserved characteristics have a very substantial effect on survival.

Table 4.5: Differences in HR between models with and without frailty in the absence of additional controls

PGM with only tariff as a covariate	HR	P-value	CI L	CI U
with frailty	0.656	<0.001	0.651	0.660
without frailty	0.684	<0.001	0.676	0.693

Heterogeneity variance=0.76. Prob.>=chibar2 < 0.001

Figure 4.5: Distribution of the Frailty

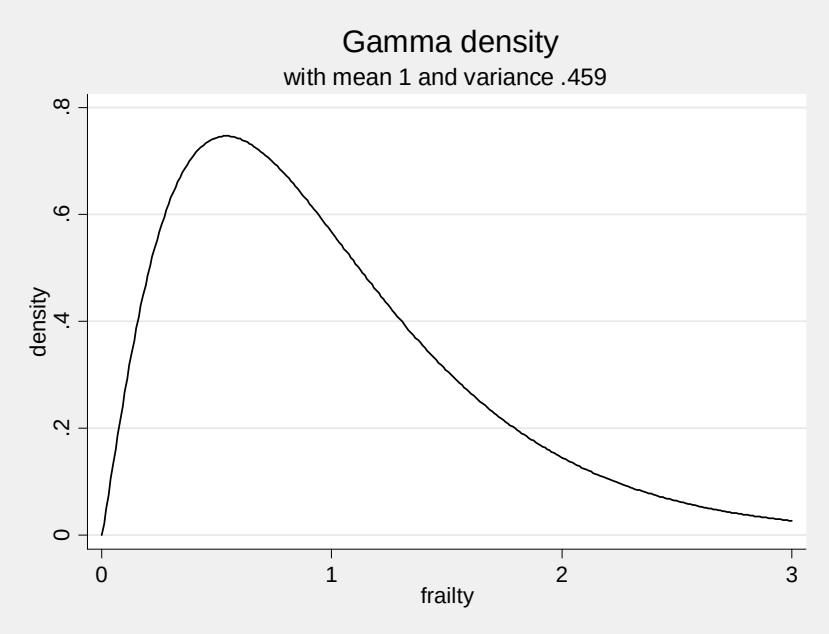


Table 4.6: Quartiles of the standard Gamma distribution

25%	Median	75%
0.502	0.852	1.338

In Figure 4.6 the impact of correction for unobserved heterogeneity on the baseline hazard in the PGM estimates are shown relatively to the model that assumes homegeneity. The correction not only has an effect on the coefficients in the regression (Table 4.7), but it also impacts on the baseline hazard, and hence implied duration dependence. As time passes the slope of the baseline hazard with heterogeneity rises faster than the hazard that assumes homogeneity.

Table 4.7: Prentice and Gloeckler Hazard model estimates (exponentiated coefficients)

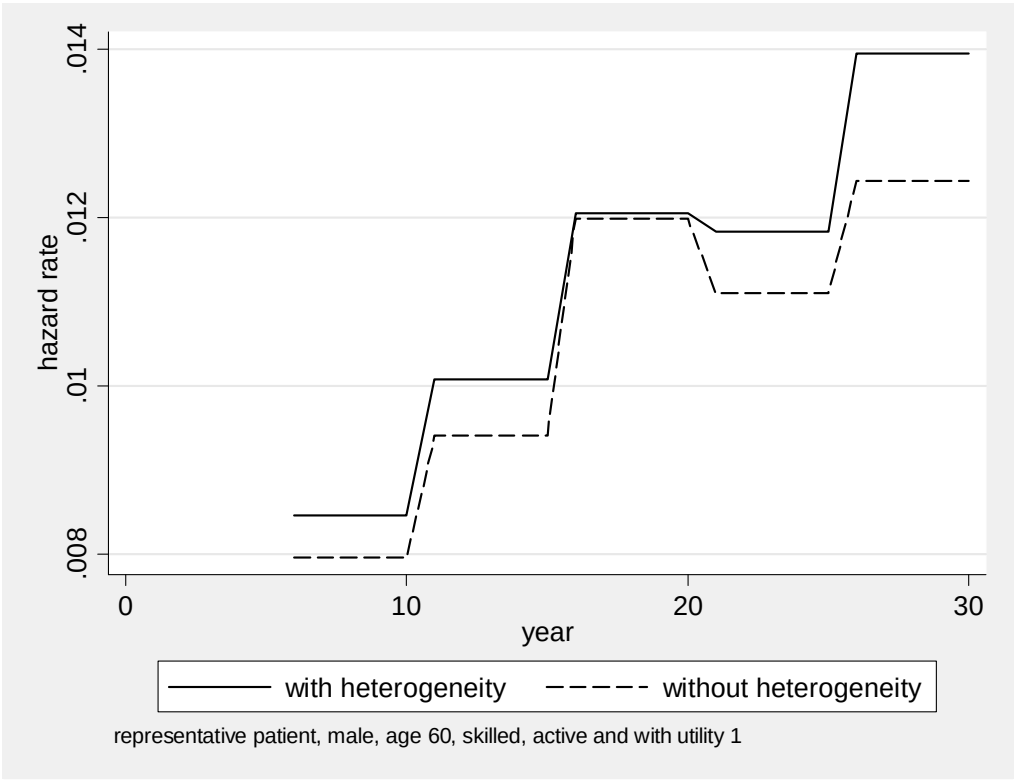
	PGM hazard model without gamma frailty				PGM hazard model with gamma frailty			
variable	**HR**	**P value**	**CI lower**	**CI upper**	**HR**	**P value**	**CI lower**	**CI upper**
EQ-5D index score per 0.10 point	0.952	<0.001	0.933	0.972	0.946	<0.001	0.925	0.969
Current Age	1.072	<0.001	1.062	1.082	1.079	<0.001	1.067	1.091
Macrovascular event	3.270	<0.001	2.492	4.290	3.684	<0.001	2.734	4.963
Microvascular event	1.527	0.083	0.946	2.467	1.600	0.072	0.959	2.670
Male(=1)	1.185	0.016	1.032	1.361	1.208	0.018	1.033	1.412
Black[a]	0.547	<0.001	0.404	0.740	0.526	<0.001	0.378	0.731
Asian[a]	0.487	<0.001	0.360	0.657	0.476	<0.001	0.345	0.656
Skilled[b]	1.085	0.320	0.924	1.274	1.116	0.232	0.932	1.336
Unskilled[b]	1.306	0.002	1.103	1.547	1.335	0.003	1.103	1.615
Smoker(=1)	1.222	0.008	1.054	1.418	1.259	0.007	1.066	1.486
Moderately active[c]	0.530	<0.001	0.463	0.607	0.500	<0.001	0.429	0.583
Active/Fit[c]	0.449	0.002	0.274	0.736	0.420	0.001	0.251	0.705
Albumin	1.101	<0.001	1.070	1.133	1.108	<0.001	1.070	1.147
Creatinin	1.108	<0.001	1.085	1.132	1.135	<0.001	1.092	1.180
t10-t14	1.168	0.335	0.852	1.600	1.178	0.303	0.862	1.610
t15-t19	1.464	0.018	1.067	2.011	1.394	0.035	1.024	1.897
t20-t24	1.420	0.054	0.994	2.029	1.268	0.162	0.909	1.770
t25-t29	1.686	0.052	0.995	2.857	1.447	0.135	0.891	2.350
σ^2					0.459	0.021		

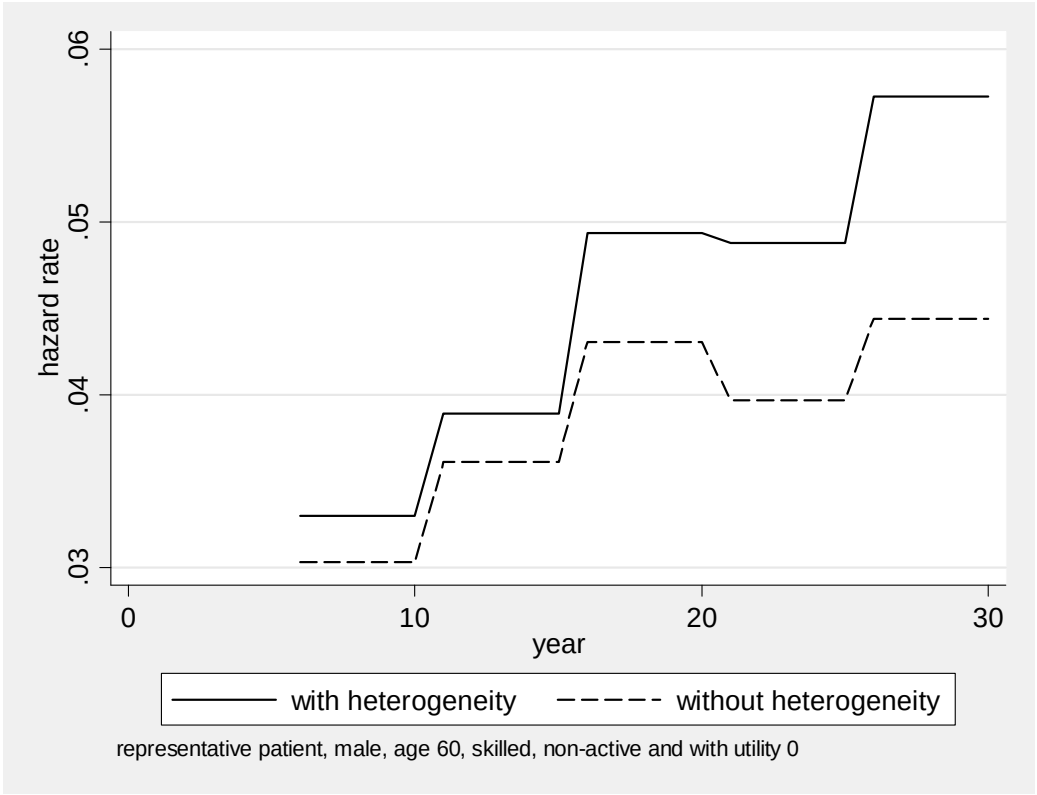
Reference category: [a]Caucasian; [b]professional and managerial, [c]Sedentary.

LR test of σ2= 0: chibar2(01) = 6.47 Prob.>=chibar2 = .005

Albumin and creatinin have been standardized as follows: $\frac{x-\bar{x}}{SD}$ where means $(\bar{x})$ and standard deviations are as those reported in Table 4.

Figure 4.6: Baseline hazard with and without heterogeneity for two sets of representative patients





N.B.:The distribution on the frailty is centred at the mean.

Predictions of life expectancy for representative patients

For predictions I assume that the error is zero and that time is modelled as a piecewise constant following five year intervals. Predictions are derived assuming that the frailty term is set equal to its mean value (=1). I first derive predicted hazard rates for persons with given characteristics, and then their implied survivor functions. The median life expectancy is compared across gender and activity levels for quartiles of the tariff distribution (e.g. tariffs=0.325, 0.71, 0.821 and 1). All other variables are set to denote the following characteristics: Caucasian, non-smoker, skilled and with no current complications. All other risk factors used to calculate h(t) are evaluated at the sample mean values, where the mean is taken over individual cases at the time of the first questionnaire, rather than over time. The representative patient is aged sixty years of age with ten years of diabetes (which corresponds to the average at the time of our first QoL measurement).

I compare predictions of median life expectancy using the coefficient estimates with and without frailty, and use partial life tables reflecting risk between 60 and 80 years. Estimates of the average number of life years are shown by level of EQ-5D index scores, gender and activity levels are shown in Tables 4.8 (with frailty) and 4.9 (without frailty).

The annual probability of mortality between the ages of 60 and 80 years is calculated using the survivor function $q = 1 - \frac{S(t+1)}{S(t)}$, where t (t=1…20) reflects number of years from an individual's 60th birthday. This probability is then adjusted for duration of diabetes. Life tables are then constructed for the sixteen types of patients considered

(male, female, active, sedentary) for each of the quartiles of the tariff's distribution. The delta method is used to estimate confidence intervals for life-expectancies.

While the differences in coefficients between the model with and without frailty were found to be small, when the difference in predicted mortality is extrapolated to life years, the risk factor compounds to as much as a one year difference in life expectancy for the average patient between the models with mean frailty and without frailty. This is quite substantial when considering an older population with a shorter available life span. Similar differences in life expectancy are achieved across gender while much bigger differences are obtained across sedentary vs. physically active individuals (up to 3 years difference in LE for the same tariff). There is no statistical difference in life expectancy across those in adjacent quartiles of the tariff distribution. Differences in survival start to be statistically significant with changes in the distribution of the tariff of more than two standard deviations. As discussed in Chapter two, clinically relevant differences in the EQ-5D index tariff are understood to be equal to one standard deviation (herein equal to 0.15).

It is interesting to note that the effects of physical activity on longevity are higher than any single decrement in the utility score. However, it is also important to point out that physical activity may be capturing some of the effects of the index score on survival. Mobility, in fact, was the most important component of the EQ-5D (Figure 4.3) to be correlated with survival.

Table 4.8: Estimated average LE (95% CI) for a person between the age of 60 to 80 years by gender, activity level, and index score level with frailty

	Active		Sedentary	
Level of Index Score	Male	Female	Male	Female
1	15.63 (15.17; 16.08)	16.76 (16.36; 17.2)	12.72 (12.14; 13.35)	13.79 (13.24; 14.37)
0.821	15.46 (15.00; 15.94)	16.39 (15.98; 16.8)	12.26 (11.68; 12.88)	13.32 (12.78; 13.90)
0.71	15.19 (14.72; 15.69)	16.15 (15.72; 16.6)	11.98 (11.40; 12.60)	13.03 (12.49; 13.61)
0.325	14.24 (13.71; 14.80)	15.26 (14.78; 15.8)	11.02 (10.41; 11.68)	12.03 (11.47; 12.63)

Table 4 9: Estimated average LE (95% CI) for a person between the age of 60 to 80 years by gender, activity level, and index score level without frailty

	Active		Sedentary	
Level of Index Score	Male	Female	Male	Female
1	16.75 (16.30; 17.28)	17.51 (17.11; 18.2)	13.78 (13.15; 14.83)	14.76 (14.18; 15.38)
0.821	16.33 (15.91; 16.82)	17.18 (16.76; 17.9)	13.33 (12.70; 14.24)	14.32 (13.74; 15.02)
0.71	16.14 (15.66; 16.65)	16.97 (16.53; 17.6)	13.05 (12.42; 13.74)	14.04 (13.46; 14.74)
0.325	15.27 (14.72; 16.06)	16.17 (15.72; 16.8)	12.08 (11.40; 13.05)	13.07 (12.45; 13.82)

4.7 Conclusion and discussion

This study looks at the link between health related quality of life (HRQoL) and survival and shows that the index scores derived from the EQ-5D are a further predictor of risk of all-cause mortality in people with Type-2 diabetes. The EQ-5D in fact appears to provide additional information about risk above and beyond that of traditional risk factors such as smoking, physical activity, and a history of non-fatal micro and macro-vascular complications. The significance and direction of the finding is in line with the published literature. I find that a 0.1 increment in index score was associated with a 5.4% (95% CI 3.1%-7.5%) lower risk of mortality. Other published studies find even more impressive impacts, ranging from 69% (Cavrini, Broccoli et al. 2012) to 8% (McEwen, Kim et al. 2009). While it is difficult to extrapolate results from different samples, I find that in the absence of covariates other than the EQ-5D index score, a 0.1 increment in index score was associated with a much higher decrement in the risk of mortality (42% when not adjusting for heterogeneity in the sample and 44% when doing so).

A number of studies have found that QoL is not only predictive of mortality but is also predictive of non-fatal events. It is important to acknowledge that this two way relation between QoL and diabetes complications could be problematic in the identification of the explanatory variables of interest. While my study is concerned purely with the direct effect of QoL on survival outcomes and hence with the correct identification of this coefficient, it should not be ruled out that when diabetes complications lead to deaths and decrements in QoL, the coefficient on complications might not be entirely identified. In order to address the concern of identification on QoL however, which is the centrepiece of my investigation, in Appendix A4.4 I test whether

or not the exclusion of diabetes complications changes in any meaningful way the coefficients on QoL. In doing so, I find that this is not the case, suggesting that the effect of non-fatal complications on QoL is much stronger than the effect of QoL on non-fatal complications.

Two innovations were explored in this chapter which to the best of my knowledge have not been tested before on any cohort of diabetic patients. The first is the use of time varying covariates, and in particular, annual measurement of HRQoL over time. The second innovation is a methodological one: to see whether or not heterogeneity may play a role in strengthening or rejecting the hypothesis of HRQoL as an independent predictor of survival.

By definition when dealing with mortality one is looking at a single spell as people only live once. One might therefore ask what is the added contribution of multiple questionnaires on QoL and other time varying covariates to the understanding of a one-off-event? Put simply, at one extreme the shape of the hazard over time could be driven solely by the changing composition of the sample under consideration, regardless of the value of the baseline covariates. To address this issue the use of time varying variables allows for the effect of covariates to be separately identified from the time pattern of the exit rate.

The methodological contribution of this chapter entails the exploration of heterogeneity in the sample, whether on observable or unobservable factors. All other studies herein summarized assumed the sample was homogeneous. However, conditional on controlling for all the important time varying risk factors, one might expect that the characteristics causing a lower hazard would be more concentrated among the remaining individuals as they approach the end of the follow-up period. For

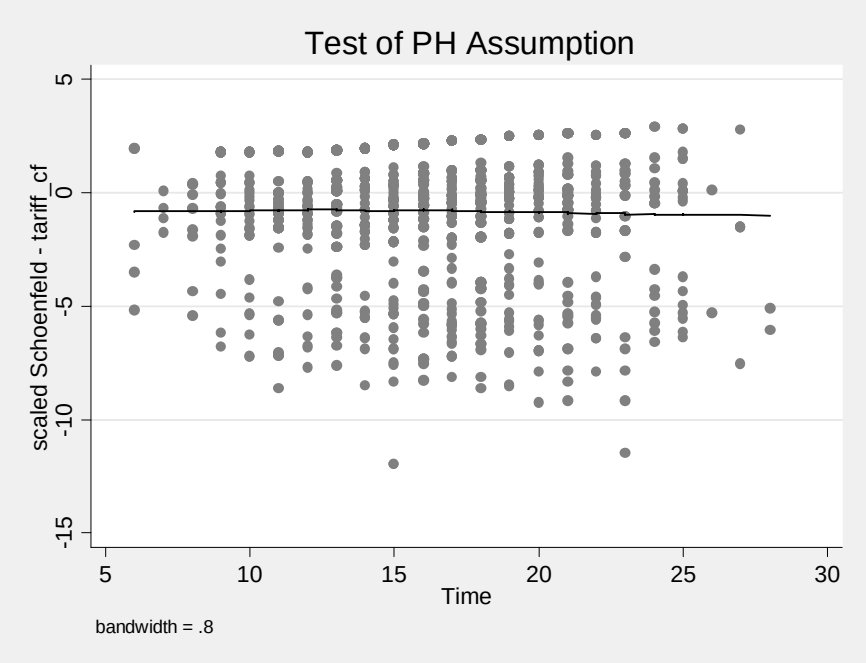
example, survivors are likely to be disproportionately represented by individuals whose genes and other characteristics lead to lower frailty. The beauty of the Prentice and Gloeckler approach (PGM) is that it is similar to Cox's partial likelihood technique (Cox (1972,1975)). Both estimators make no assumptions about the baseline hazard and the Cox proportional hazard model is asymptotically equivalent to the Prentice and Gloeckler approach. Appendix A4.3 shows this equivalence.

There are a number of potential extensions to this work. Following the concerns about identification outlined in Appendix A4.4, an interesting possible extension would involve an exploration of the predictive power of the index scores towards multiple spells of different non-fatal complications. Given that this is the first study employing multiple QoL questionnaires, another possible extension would be to determine the optimal time between different rounds of questionnaire administration so as to understand the duration over which index scores can predict elevations in risk. It should be noted that a possible consequence of carrying forward the values of the questionnaire in years when we have non-response may have lead to conservative estimates of the impact of self-reported health on mortality given that QoL is likely to decrease over time and that QoL is negatively correlated with the probability of death. Additionally, when interpreting the results herein reported we ought to bear in mind possible issues surrounding the external validity of the findings as these are subject to the data available and the characteristics of the trial and its participants. Similarly, the frailty in the sample was modelled using a convenient parametric distribution. If one were interested in different moments of the distribution besides the mean, the magnitude of the impact of the unobserved heterogeneity on the probability of survival is likely to differ with the distributional assumptions adopted.

The implications of this work are twofold. First, as many authors have previously stated, EQ-5D index scores could have value in a clinical setting in the identification of high risk patients since those with lower index scores are also likely to have also a lower life expectancy. More importantly, however, heterogeneity matters and is duration dependent.

Appendix

A4. 1: Test of the proportional-hazards assumption for Tariff



A4. 2: Non-Exponentiated Prentice and Gloeckler Results

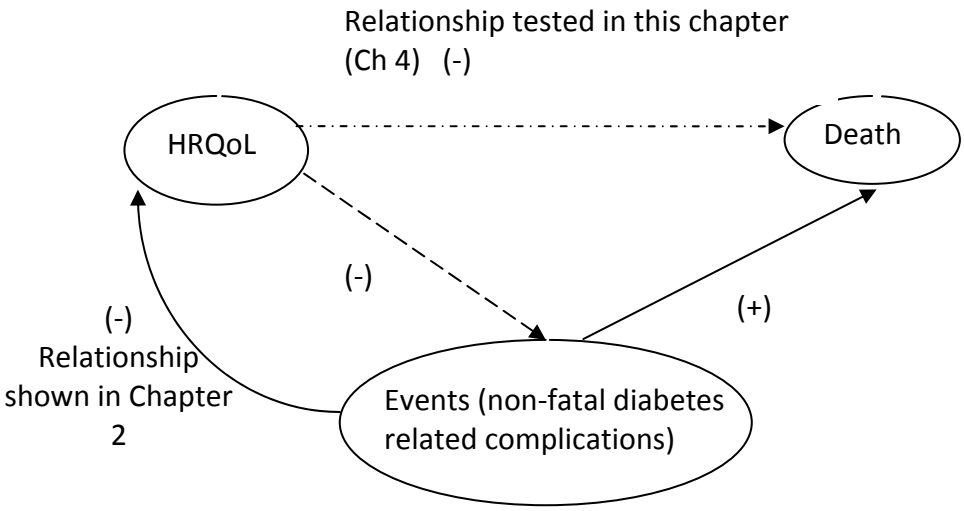
variable	PGM hazard model without gamma frailty		PGM hazard model with gamma frailty	
	Coef.	Std. Err.	Coef.	Std. Err.
EQ-5D index score per 0.10 point	-0.049**	(0.011)	-0.055**	(0.012)
Current Age	0.069**	(0.005)	0.076**	(0.006)
Macrovascular event	1.185**	(0.139)	1.304**	(0.152)
Microvascular event	0.423	(0.245)	0.470	(0.261)
Male(=1)	0.170*	(0.071)	0.189*	(0.080)
Black*	-0.604**	(0.154)	-0.642**	(0.168)
Asian*	-0.721**	(0.153)	-0.742**	(0.164)
Skilled**	0.081	(0.082)	0.110	(0.092)
Unskilled**	0.267**	(0.086)	0.289**	(0.097)
Smoker(=1)	0.201**	(0.076)	0.230**	(0.085)
Moderately active***	-0.635**	(0.069)	-0.692**	(0.078)
Active/Fit***	-0.801**	(0.252)	-0.867**	(0.264)
Albumin	0.096**	(0.015)	0.102**	(0.018)
Creatinin	0.103**	(0.011)	0.127**	(0.020)
t10-t14	0.164	(0.159)	0.155	(0.161)
t15-t19	0.332*	(0.157)	0.381*	(0.162)
t20-t24	0.238	(0.170)	0.351	(0.182)
t25-t29	0.370	(0.247)	0.522	(0.269)
Constant	-8.076**	(0.354)	-8.475**	(0.416)

A4.3: Comparing Prentice and Gloeckler with and without frailty with the Cox PH model

	Cox PH Model				PGM hazard model without gamma frailty				PGM hazard model with gamma frailty			
variable	**HR**	**P value**	**CI lower**	**CI upper**	**HR**	**P value**	**CI lower**	**CI upper**	**HR**	**P value**	**CI lower**	**CI upper**
EQ-5D index score per 0.10 point	0.954	<0.001	0.935	0.974	0.952	<0.001	0.933	0.972	0.946	<0.001	0.925	0.969
Current Age	1.070	<0.001	1.060	1.079	1.072	<0.001	1.062	1.082	1.079	<0.001	1.067	1.091
Macrovascular event	3.019	<0.001	2.300	3.963	3.270	<0.001	2.492	4.290	3.684	<0.001	2.734	4.963
Microvascular event	1.375	0.194	0.850	2.223	1.527	0.083	0.946	2.467	1.600	0.072	0.959	2.670
Male(=1)	1.189	0.014	1.035	1.365	1.185	0.016	1.032	1.361	1.208	0.018	1.033	1.412
Black	0.562	<0.001	0.416	0.760	0.547	<0.001	0.404	0.740	0.526	<0.001	0.378	0.731
Asian	0.497	<0.001	0.368	0.671	0.487	<0.001	0.360	0.657	0.476	<0.001	0.345	0.656
Skilled	1.084	0.327	0.923	1.272	1.085	0.32	0.924	1.274	1.116	0.232	0.932	1.336
Unskilled	1.295	0.003	1.094	1.534	1.306	0.002	1.103	1.547	1.335	0.003	1.103	1.615
Smoker(=1)	1.212	0.011	1.045	1.406	1.222	0.008	1.054	1.418	1.259	0.007	1.066	1.486
Moderately active	0.544	<0.001	0.475	0.623	0.530	<0.001	0.463	0.607	0.500	<0.001	0.429	0.583
Active/Fit	0.463	0.002	0.282	0.760	0.449	0.002	0.274	0.736	0.420	0.001	0.251	0.705
Albumin	1.097	<0.001	1.065	1.129	1.101	<0.001	1.070	1.133	1.108	<0.001	1.070	1.147
Creatinin	1.102	<0.001	1.078	1.126	1.108	<0.001	1.085	1.132	1.135	<0.001	1.092	1.180

The three specifications above show similar results. However, both the Cox PH model and the PGM model without frailty show a lower impact of the tariff than the specification that takes into account heterogeneity. It is important to point out that the Prentice and Gloeckler approach (PGM) is similar to Cox's partial likelihood technique (Cox (1972,1975)). Both estimators make no assumptions about the baseline hazard. Bailey (1984) shows that the Cox PH model is asymptotically equivalent to the Prentice and Gloeckler approach.

A4.4: Can HRQoL lead to non-fatal diabetes-related complications?



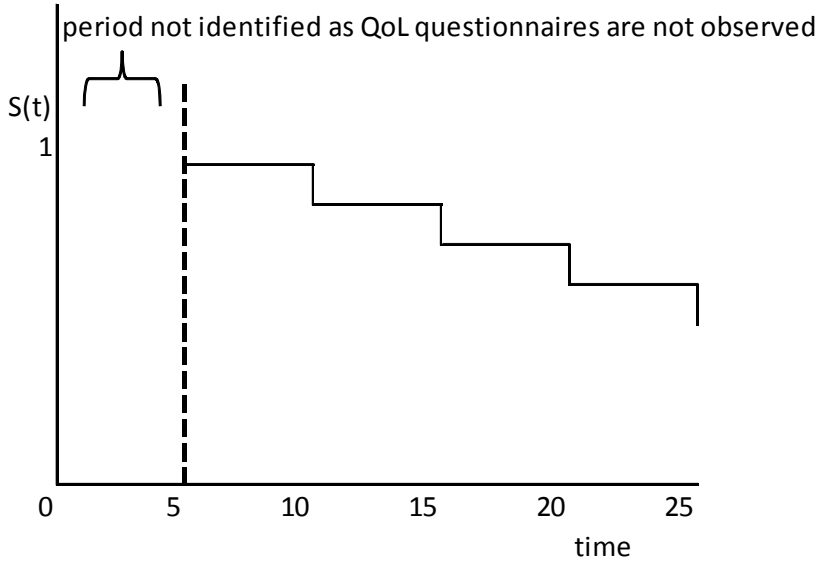
Above is a directed acyclic graph showing the possible relations between some key variables of interest. If complications affect both QoL (as shown in chapter 2), including complication is important if complications also affect survival, which is well documented in the medical literature. It is possible, however, that HRQoL also affects the likelihood of non-fatal events, in which case the causal impact of events on survival will not be well identified as QoL would enter the causal path. Both Clarke et al. (2009) and Hayes et al. (2008), using the FIELD dataset, found that QoL is also a predictor of first time macro-vascular events. If the direction of causality were only from complications to QoL, leaving complications out of the model would provide the total effect of QoL on mortality. In order to see if including complications might affect the coefficient on QoL, I reran the model in Table 4.5, but omitting both micro and macro vascular complications. The impact of QoL on survival was left unchanged suggesting that the most likely and more powerful relationship is from diabetes complications to QoL rather than the other way around. It is important to point out that this may affect the identification of the coefficient on the diabetes-related

complications. The aim of this study, however, is to determine whether QoL has a direct effect on survival outcomes conditional on the events it might cause or is influenced by.

Re-estimation of Table 4.5 without micro and macro-vascular complications

	PGM hazard model without gamma frailty				PGM hazard model with gamma frailty			
variable	**HR**	**P value**	**CI lower**	**CI upper**	**HR**	**P value**	**CI lower**	**CI upper**
EQ-5D index score per 0.10 point	0.951	<0.001	0.931	0.970	0.947	<0.001	0.928	0.969
Current Age	1.072	<0.001	1.063	1.082	1.074	<0.001	1.063	1.086
Male(=1)	1.200	0.01	1.045	1.378	1.209	0.01	1.046	1.397
Black	0.534	<0.001	0.395	0.723	0.527	<0.001	0.386	0.720
Asian	0.489	<0.001	0.362	0.660	0.486	<0.001	0.358	0.660
Skilled	1.085	0.318	0.924	1.274	1.094	0.292	0.926	1.293
Unskilled	1.305	0.002	1.102	1.546	1.314	0.002	1.102	1.568
Smoker(=1)	1.222	0.008	1.053	1.417	1.233	0.008	1.056	1.439
Moderately active	0.520	<0.001	0.455	0.596	0.511	<0.001	0.442	0.591
Active/Fit	0.442	0.001	0.270	0.726	0.434	0.001	0.262	0.717
Albumin	1.100	<0.001	1.069	1.132	1.103	<0.001	1.069	1.137
Creatinin	1.108	<0.001	1.084	1.132	1.115	<0.001	1.082	1.149
t5-t9	0.000	<0.001	0.000	0.001	0.000	<0.001	0.000	0.001
t10-t14	0.000	<0.001	0.000	0.001	0.000	<0.001	0.000	0.001
t15-t19	0.000	<0.001	0.000	0.001	0.000	<0.001	0.000	0.001
t20-t24	0.000	<0.001	0.000	0.001	0.000	<0.001	0.000	0.001
t25-t29	0.000	<0.001	0.000	0.001	0.000	<0.001	0.000	0.001

A4. 5: LE prediction strategy



I consider a period of 20 years in order for results to be comparable to other publications, in particular Clarke et al. (2009).

The representative patient is assumed to be 60 years old at the time of questionnaire 1. This means that the patient is 55 years old at the time of diagnosis. The life expectancy is calculated as follows:

$$LE = \sum_{t=5}^{25} \frac{\prod_{s=1}^{t} S(s) + \prod_{s=1}^{t-1} S(s)}{2}$$

where S(s) varies according to the five-year intervals used to specify the baseline hazard.

References

Aalen, O. O. (1988). "Heterogeneity in survival analysis." Statistics in Medicine **7**(11): 1121-1137.

Bailey, K. R. (1984). "Asymptotic Equivalence Between the Cox Estimator and the General ML Estimators of Regression and Survival Parameters in the Cox Model." The Annals of Statistics **12**(2): 730-736.

Cavrini, G., S. Broccoli, et al. (2012). "EQ-5D as a predictor of mortality and hospitalization in elderly people." Quality of Life Research **21**(2): 269-280.

Clarke, P. M., A. J. Hayes, et al. (2009). "Using the EQ-5D Index Score as a Predictor of Outcomes in Patients With Type 2 Diabetes." Medical Care **47**(1): 61-68 10.1097/MLR.0b013e3181844855.

Cleves, M., W. Gould, et al. (2010). An Introduction to Survival Analysis Using Stata. College Station, Texas, StataCorp.

Fiscella, K. and P. Franks (1997). "Poverty or income inequality as predictor of mortality: longitudinal cohort study." BMJ **314**(7096): 1724.

Gutierrez, R., G. S. Carter, et al. (2001). "On boundary-value likelihood-ratio tests." Stata Technical Bulletin **10**: 269-273.

Hayes, A. J., P. M. Clarke, et al. (2008). "Can Self-Rated Health Scores Be Used for Risk Prediction in Patients With Type 2 Diabetes?" Diabetes Care **31**(4): 795-797.

Hayes, A. J., P. M. Clarke, et al. (2011). "Simulation of Quality-Adjusted Survival in Chronic Diseases: An Application in Type 2 Diabetes." Medical Decision Making.

Heckman, J. and B. Singer (1984). "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." Econometrica **52**(2): 271-320.

Idler, E. L. and Y. Benyamini (1997). "Self-rated health and mortality: a review of twenty-seven community studies." Journal of health and social behavior: 21-37.

Jenkins, S. P. (2008). "Unobserved heterogeneity." Survival Analysis using Stata **Chapter 7**.

Jerant, A., D. J. Tancredi, et al. (2011). "Mortality Prediction by Quality-Adjusted Life Year Compatible Health Measures: Findings in a Nationally Representative US Sample." Medical Care **49**(5): 443-450 10.1097/MLR.0b013e318206c231.

Kiecolt-Glaser, J. K. and R. Glaser (2002). "Depression and immune function-Central pathways to morbidity and mortality." Journal of Psychosomatic Research **53**(4): 873-876.

Kleefstra, N., G. W. D. Landman, et al. (2008). "Prediction of Mortality in Type 2 Diabetes From Health-Related Quality of Life (ZODIAC-4)." Diabetes Care **31**(5): 932-933.

Lancaster, T. (1990). The Econometric Analysis of Transition Data
Cambridge, Cambridge University Press.

Landman, G. W. D., K. J. J. van Hateren, et al. (2010). "Health-Related Quality of Life and Mortality in a General and Elderly Population of Patients With Type 2 Diabetes (ZODIAC-18)." Diabetes Care **33**(11): 2378-2382.

Lenzen, M. J., W. J. S. Reimer, et al. (2007). "The additional value of patient-reported health status in predicting 1-year mortality after invasive coronary procedures: a report from the Euro Heart Survey on Coronary Revascularisation." Heart **93**(3): 339-344.

Lin, D. Y. and L. J. Wei (1989). "The Robust Inference for the Cox Proportional Hazards Model." Journal of the American Statistical Association **84**(408): 1074-1078.

McEwen, L. N., C. Kim, et al. (2009). "Are health-related quality-of-life and self-rated health associated with mortality? Insights from Translating Research Into Action for Diabetes (TRIAD)." Primary care diabetes **3**(1): 37-42.

McEwen L.N., Kim C. et al. (2006). TRIAD Study Group. “Diabetes reporting as a cause of death: results from the Translating Research into Action for Diabetes (TRIAD) study”. Diabetes Care: **29**:247–253.

Meyer, B. D. (1990). "Unemployment insurance and unemployment spells." Econometrica **58**(4): 757-782.

Park, S. M., M. H. Park, et al. (2006). "EuroQol and survival prediction in terminal cancer patients: a multicenter prospective study in hospice-palliative care units." Supportive Care in Cancer **14**(4): 329-333.

Prentice, R. and L. Gloeckler (1978). "Regression analysis of grouped survival data with application to breast cancer data." Biometrics **34**: 57-67.

Prentice, R. L. and G. L. A. (1978). "Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data." Biometrics **34**(1): 57-67.

Royston, P. and P. C. Lambert (2011). Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Mode. College Station, StataCorp.

Schoenfeld, D. (1982). "Partial residuals for the proportional hazards regression model." Biometrika **69**(1): 239-241.

StataCorp, 2011, "Stata 12 Base Reference Manual, College Station", TX, Press.

Stratton, I. M., A. I. Adler, et al. (2000). "Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study." BMJ **321**(7258): 405-412.

Trussell, J. and T. Richards (1985). Correcting for Unmeasured Heterogeneity in Hazard Models Using the Heckman-Singer Procedure. San Francisco, Josses-Bass.

Vaupel, J. W., K. G. Manton, et al. (1979). "The Impact of Heterogeneity in Individual frailty on the Dynamics of Mortality." Demography **16**: 439-454.

Wennberg, P., O. Rolandsson, et al. (2012). "Self-rated health and mortality in individuals with diabetes mellitus: prospective cohort study." BMJ Open **2**(1).

Yamaguchi, K. (1986). Alternative Approaches to Unobserved Heterogeneity in the Analysis of Repeatable Events. Sociological Methodology. N. Tuma. San Francisco, Jossey-Bass**:** 213-49.

Chapter 5

CONCLUSION

5.1 Review of the thesis

The work of this thesis is concerned with the impact of type 2 diabetes-related complications on utility and healthcare costs. It also looks at whether self-reported health related quality of life can be an independent predictor of mortality in diabetes.

This is an important area of research as treatments available to reduce complications of diabetes are rapidly expanding (Alexander G. et al., 2008). In order to understand the economic benefits of investment in such therapies we need estimates of the impacts of diabetes-related complications on patients' quality of life and healthcare costs as well as the impact of quality of life on survival.

This project has benefitted greatly from access to an important data set which was the product of thirty laborious years of research, and in particular from the work that went into designing the instruments (HRQoL and resource use questionnaires) herein used. Multiple questionnaires were administered to patients in order to improve on the estimates previously published using a cross section of the UKPDS. This has allowed me to study the importance of patient heterogeneity.

The contribution of the thesis concerns the exploration of a range of issues to answer reliably the questions of interest, including appropriate panel data techniques, the importance of functional form issues, dealing with attrition over time, and taking into account the effects of selection and clustering. A unifying theme across the thesis is

the challenge of estimating causal parameters in the context of substantial observed and unobserved patient heterogeneity.

Many of the methodological and empirical features of this thesis are new to this literature, such as the use of multiple questionnaires, the exploration of within patient variation for the outcomes of interest, and the use of time varying covariates and frailty in survival. If this is important and new, why didn't anybody do it before? Part of the reason may be that economic evaluations still tend to be seen as adjuncts to clinical trials. They are not the main purpose for which the trials are set up, and consequently data that would be valuable for economic analysis may not be prioritised. For instance, I use data from several rounds of questionnaires, but this is the first published analysis of patients participating in a large diabetes trial who were administered the EQ5D questionnaire more than twice. Similarly, the resource use questionnaires administered on multiple occasions in this study are seldom used in this way.

The main findings of this DPhil research are summarised in Section 5.2, together with some practical implications and comparisons to other recent applied studies in the field. Section 5.3 outlines the limitations and generalisations of this work.

5.2 Findings and implications

This section aims to answer the following questions for each one of the substantive chapters of this thesis: Why are the topics of this study important? Why do we want or need to evaluate the impact of diabetes complications on QoL and costs? Or for that matter any other disease? How do the results of this work compare with the existing literature and what are the implications of my findings?

CHAPTER 2 – the impact of diabetes-related complications on quality of life

Importance of the research question

In order to understand the burden of a disease, we have two primary alternatives. One approach is to run a standard gamble or time-trade-off study on patients or the general public where the aim is to collapse the quality and length of life into a single measure. In practice there are limited opportunities for conducting such studies. The other alternative is to use reference values for the utility associated with particular health states. Typical sources of information are population surveys providing normative data on usual quality-of-life levels.

Herein I link answers from a self-reported multiple attribute questionnaire to weights derived from a population based TTO study in order to understand the burden of the disease and produce preference-based QoL measures. Cost-utility analysis requires preference-based measures of health-related QoL in order to estimate utility values of competing treatments.

This brings me to the other reason why having estimates of the impact of complications on QoL is important, which is precisely in the context of cost effectiveness work. We are rapidly expanding treatments available to reduce complications of diabetes. Consequently, in order to understand the economic benefits of investment in such therapies we need estimates of the impacts of diabetes-related complications on patients' quality of life. Because patient heterogeneity accounts for much of the variation in quality of life, estimates that ignore such heterogeneity can be severely biased.

Summary of the findings

Patients who experience diabetes-related complications appear to be systematically different from those who do not; not only in terms of observable characteristics but also in terms of unobservable heterogeneity. In particular, such patients have a tendency to experience a lower quality of life even prior to the onset of complications. In the sample of UKPDS patients answering HRQoL questionnaires, if two patients are drawn randomly from the sample, the difference between their QoL levels is nearly twice as large as the difference for the same patient over two randomly selected years. This suggests there is considerable across-patient heterogeneity and persistence in QoL levels. The results showed that amputation (-0.172 SE[0.029]) and stroke (-0.165 SE[0.022]) have the largest impact on quality of life, amounting to a decrement of slightly more than a half standard deviation in the utility scale used. To put these results into perspective, the bar for a clinically relevant difference suggested by Norman and co-authors is about half a standard deviation for chronically ill patients. On this basis, the impact of heart failure and myocardial infarction is statistically significant but is not clinically relevant. We do not find any statistically significant effect of blindness in one eye on quality of life (0.033 SE[0.023]) or of ischemic heart disease (-0.029 SE[0.022]). These results are very different both when compared to earlier published cross sectional results on the same cohort and when compared to the cross sectional results on the same data set. The differences in coefficients are close to half a standard deviation of the distribution of quality of life.

Implications

This chapter highlights the importance of studying quality of life changes over time in order to distinguish between time-invariant determinants of QoL and the effect on QoL of specific events such as diabetes-related complications. This, in practice, translates into a call for making the collection of longitudinal data standard practice. At the policy level this conclusion is also relevant to health systems less heavily reliant than the NHS on cost-effectiveness analysis in determining treatment priorities, because directly or indirectly all systems face budget constraints when deciding to invest in the development of new interventions.

The clinical literature has often used a cross-sectional approach to attribute differences in quality of life between patients with and without events, cross-sectional data being less expensive and easier to collect than longitudinal data. In my work I show instead that differences are largely due to underlying and potentially unobserved heterogeneity across these two groups of patients. I show this by applying econometric techniques that take into account attrition and FE using a unique dataset that tracks patients from the UK prospective diabetes study.

The policy implications of my work are twofold. In this particular case cross sectional results have been shown to overestimate the impact of complications on QoL, so compared with previous cross sectional estimates the threshold to introduce new diabetes related interventions should be higher than it currently is. Secondly, since panel methods address an important source of bias, researchers who are interested in the economic consequences of chronic conditions should be collecting longitudinal data.

CHAPTER 3 – impact of diabetes-related complications on health care costs

Importance of the research question

There are close to 3 million patients with diabetes in the UK alone (Merrick D., 2005) and almost 300 million worldwide (Wild, Roglic et al., 2004). Many more have undiagnosed diabetes. A wide range of complications are associated with diabetes, making it important to get accurate and unbiased estimates of the impact of these complications on inpatient and non-inpatient costs in the short and long term.

Summary of the findings

The results reported in this chapter indicate the substantial impact of many diabetes-related complications on both hospital costs and non-inpatient expenditures, not only in the year in which an event occurs, but in permanently raising the average level of these in all subsequent years. The largest average annual costs are attributable to amputation in both inpatient and non-inpatient settings, followed by ischemic heart disease, myocardial infarctions and stroke.

I also find that individuals' propensity to use healthcare resources does not change much over time with respect to GP visits and other non-inpatient related expenditures, in contrast to hospitalizations where levels of use rise with the patient's age and time since diagnosis of diabetes. One interpretation of these results is that a subset of patients routinely uses non-inpatient resources whereas inpatient resources, which are not elective occur as though they were randomly assigned across individuals. In fact, in the case of inpatient costs there is no evidence of FE, while in the case of non-inpatient utilization the hypothesis of a common intercept is rejected. This is perhaps intuitive insofar as hospitalizations are dominated by non-elective cases, whereas in the

case of GP visits, there a stronger component of patient-specific choice and preferences and so different underlying levels of utilization are present across patients, irrespective of the likelihood of events. I also find that within-patient differences lead to slightly lower marginal effects of diabetes complications on costs than between-patient differences, as patients with diabetes-related events are on a higher cost path at least two years before a complication occurs.

Implications

Non-inpatient care consists of visits to general practitioners, contact hours with nurses and other health care practitioners, and remains the norm for most routine care. While non-inpatient resource utilization represents roughly 30% of total expenditure, the proportion of costs attributable to non-inpatient care has steadily risen over time and this corresponds with the publication of guidance protocols demonstrating the benefits of intensive monitoring and prevention. In contrast, hospitalizations are mainly associated with acute events, and occur for major and often unplanned diagnostic, surgical or therapeutic services. The implication of this chapter is that unobserved patient heterogeneity might be important in non-inpatient care utilization while it is less relevant in inpatient care.

CHAPTER 4 – Self-reported quality of life as a predictor of survival

Importance of the research question

It is important to have accurate measures of the impact of QoL on the risk of mortality especially as these two components are combined into quality adjusted life years (QALYs) which is then used to allocate healthcare resources across competing interventions. Standard survival models often assume homogeneity: all individuals are subject to the same risks embodied in the hazard (t) or the survivor functions S(t). Over time, however, even if unobservable variables are orthogonal to covariates, the initial distribution of patients in the study differs from final distribution of survivors in the sample, which by construction are less frail.

Summary of the findings

The methodological contribution of this chapter entails the exploration of unobserved heterogeneity in the sample. Conditional on controlling for all the important time varying risk factors, lower frailty is concentrated amongst survivors. Thus over time the slope of the individual hazard rises faster than the population hazard. Accounting for heterogeneity translates into a stronger effect of QoL on survival. While the differences in coefficients between the model with and without frailty were found to be small, when the difference in predicted mortality is extrapolated to life years, the difference in risk compounds to as much as a one-year difference in life expectancy for the average patient between the model with mean frailty and that without frailty. The variance of frailty across individuals however is such that the average difference in hazards across questionnaire responders is 45.9%.

Implications

HRQoL questionnaires could be used to identify high risk patients to target in the management of diabetes care, and could complement physical health assessments. The EQ-5D in particular is an inexpensive and easy-to-administer summary measure and appears to be a proxy for other indicators that may be more difficult to collect.

Accurate estimates of the relationship between HRQoL and survival is also important when extrapolation of both life expectancy and the HRQoL profile is required. This is relevant when QALYs are required to quantify health outcomes in cost-utility analysis. Since health is a function of length of life and quality of life and the QALY combines the value of these attributes into a single index number, if patients with lower health state utilities are at higher risk of mortality, this would impact on their expected future QALYs.

5.3 Limitations and Generalizability

The strength of this study is the careful consideration of the ways in which patient specific heterogeneity can lead to bias either through selection into the sample or because of unobserved patient-specific characteristics correlated with the variables of interest. The weakness of the findings is the external validity of the results. It should be acknowledged that UKPDS patients are possibly not representative of other people with diabetes as they were diagnosed in the 1970s and 80s (the age of diagnosis for instances has decreased considerably since then) and were part of a clinical trial. The large number of patients recruited in the UKPDS provides a certain assurance that similar results should apply to other populations. Self-selection into trials is however often unavoidable and the unintended cause of this is that certain groups (notably women and ethnic minorities) tend to be underrepresented despite the effort to achieve a

balanced demographic representation. Women in the UKPDS for instance are slightly under-represented while the ethnic composition of the UKPDS patients is in line with the UK ethnic composition at the time of randomization.

More work always remains to be done, but I hope this work and results will be of interest and use to other researchers, and will contribute to future provision of cost-effective care to the growing number of people with type 2 diabetes around the world.

References

Alexander G, Sehgal NL, Moloney RM et al. (2008). "National Trends in Treatment of Type 2 Diabetes Mellitus, 1994-2007". Arch Intern Med.:**168**(19):2088-2094.

Merrick D. (2005). "The PBS diabetes population prevalence model", Heslington, UK: Yorkshire and Humber Public Health Observatory, University of York.

Wild S, Roglic G, et al. (2004). "Global prevalence of diabetes: Estimates for the year 2000 and projections for 2030." Diabetes Care **27**: 1047-1053.